Europäisches Patentamt | **European Patent Office** | **Office européen des brevets**

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. **Patent application No.** **Demande de brevet n°**

98202467.1

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 98202467.1
Demande no:

Anmeldetag:
Date of filing: 22.07.98
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Stichting Instituut voor Dierhouderij
en Diergezondheid (ID-DLO)
Edelhertweg 15
8219 PH Lelystad
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Streptococcus suis vaccines and diagnostic tests

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/54
C12N9/10
C12N1/21
C07K14/315
C12P19/04
A61K39/09
C12R 146

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

EPO - DG 1

Eur 4573

22. 07. 1998

Title: *Streptococcus suis* vaccines and diagnostic tests.

The invention relates to *Streptococcus suis* infections of pigs, to vaccines directed against those infections and to tests for diagnosing *Streptococcus suis* infections.

Streptococcus suis is an important cause of meningitis, septicemia, arthritis and sudden death in young pigs (4, 46). Incidentally, it can also cause meningitis in man (1). *S.suis* strains are usually identified and classified by their morphological, biochemical and serological characteristics (58, 59, 46). Serological classification is based on the presence of specific antigenic polysaccharides. So far, 35 different serotypes have been described (9, 56, 14). In several European countries, *S. suis* serotype 2 is the most prevalent type isolated from diseased pigs, followed by serotypes 9 and 1. Serological typing of *S. suis* is carried out using different types of agglutination tests. In these tests, isolated and biochemically characterised *S. suis* cells are agglutinated with a panel of 35 specific sera. These methods are very laborious and time-consuming.

Little is known about the pathogenesis of the disease caused by *S. suis,* let alone about its various serotypes such as type 2. Various bacterial components, such as extracellular and cell-membrane associated proteins, fimbriae, haemaglutinins, and haemolysin have been suggested as virulence factors (9, 10, 11, 15, 16, 47, 49). However, the precise role of these protein components in the pathogenesis of the disease remains unclear (37). It is well known that the polysaccharidic capsule of various Streptococci and other gram-positive bacteria plays an important role in pathogenesis (3, 6, 35, 51, 52). The capsule enables these micro-organisms to resist phagocytosis and is therefore regarded as an important virulence factor. Recently, a role of the capsule of *S. suis* in the pathogenesis was suggested as well (5). However, the structure, organisation and

functioning of the genes responsible for capsule polysaccharide synthesis (*cps*) in *S. suis* is unknown. Within *S. suis* serotypes 1 and 2 strains can differ in virulence for pigs (41, 45, 49). Some type 1 and 2 strains are virulent, other strains are not. Because both virulent and non-virulent strains of serotype 1 and 2 strains are fully encapsulated, it may even be that capsule is not a relevant factor required for virulence.

Attempts to control *S. suis* infections or disease are still hampered by the lack of knowledge about the epidemiology of the disease and the lack of effective vaccines and sensitive diagnostics.

The invention provides an isolated or recombinant nucleic acid encoding a capsular (*cps*) gene cluster of *Streptococcus suis*. Biosynthesis of capsule polysaccharides in general has been studied in a number of Gram-positive and Gram-negative bacteria (32). In Gram-negative bacteria, but also in a number of gram-positive bacteria, genes which are involved in the biosynthesis of polysaccharides are clustered at a single locus. *Streptococcus suis* capsular genes as provided by the invention show a common genetic organisation involving three distinct regions. The central region is serotype specific and encodes enzymes responsible for the synthesis and polymerisation of the polysaccharides. This region is flanked by two regions conserved in *Streptococcus suis* which encode proteins for common functions such as transport of the polysaccharide across the cellular membrane. However, in between species, only low homologies exist, hampering easy comparison and detection of seemingly similar genes. Knowing the nucleic acid encoding the flanking regions allows type-specific determination of nucleic acid of the central region of *Streptococcus suis* serotypes, as for example described in the experimental part of the description of the invention.

The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derived thereof. Such a nucleic acid

is for example provided by hybridising chromosomal DNA derived from any one of the *Streptococcus suis* serotypes to a nucleic acid encoding a gene derived from a *Streptococcus suis* serotype 1, 2 or 9 capsular gene cluster, as provided by the invention (see for example Tables 4 and 5) and cloning of (type-specific) genes as for example described in the experimental part of the description. At least 14 open reading frames are identified. Most of the genes belong to a single transcriptional unit, identifying a co-ordinate control of these genes, they, and the enzymes and proteins they encode, act in concert to provide the capsule with the relevant polysaccharides. The invention provides *cps* genes and proteins encoded thereof involved in regulation (CpsA), chain length determination (CpsB, C), export (CpsC) and biosynthesis (CpsE, F, G, H, J, K). Although the overall organisation seemed at first glance to be similar to that of the *cps* and *eps* gene clusters of a number of Gram-positive bacteria (19, 32, 42), overall homologies are low (see table 3). The region involved in biosynthesis is located at the centre of the gene cluster and is flanked by two regions containing genes with more common functions.

The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 2 or a gene or gene fragment derived thereof, preferably as identified in Figure 3. Genes in this gene cluster are involved in polysaccharide biosynthesis of capsular components and antigens. For a further description of such genes see for example Table 2 of the description, for example a cpsA gene is provided functionally encoding regulation of capsular polysaccharide synthesis, whereas cpsB and cpsC are functionally involved in chain in chain length determination. Other genes, such as cpsD, E, F, G, H, I, J, K and related genes, are involved in polysaccharide syntheses, functioning for example as glucosyl- or glycosyltransferase. The cpsF, G, H, I, J genes encode more type-specific proteins than the flanking genes which are found more-or-less conserved

throughout the species and can serve as base for selection of primers or probes in PCR-amplification or cross-hybridisation experiments for subsequent cloning.

For example, the invention further provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 1 or a gene or gene fragment derived thereof, preferably as identified in Figure 4.

In addition, the invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 9 or a gene or gene fragment derived thereof, preferably as identified in Figure 5.

Furthermore, the invention provides for example a fragment or parts thereof of the *cps* locus, involved in the capsular polysaccharide biosynthesis, of *S. suis*, exemplified in the experimental part for serotype 1, 2 or 9, and allows easy identification or detection of related fragments derived of other serotype of *S. suis*.

The invention provides a nucleic acid probe or primer derived from a nucleic acid according to the invention allowing species or serotype specific detection of *Streptococcus suis*. Such a probe or primer (herein used interchangeably) is for example a DNA, RNA or PNA (peptide nucleic acid) probe hybridising with capsular nucleic acid as provided by the invention. Species specific detection is provided preferably by selecting a probe or primer sequence from a species-specific region (e.g. flanking region) whereas serotype specific detection is provided preferably by selecting a probe or primer sequence from a type-specific region (e.g. central region) of a capsular gene cluster as provided by the invention. Such a probe or primer can be used in a further unmodified form, for example in cross-hybridisation or polymerase-chain reaction (PCR) experiments as for example described in the experimental part of the description of the invention. Herein the invention provides the isolation and molecular characterisation of additional

type specific *cps* genes of *S. suis* types 1 and 9. In addition, we describe the genetic diversity of the *cps* loci of serotypes 1, 2 and 9 among the 35 *S. suis* serotypes yet known. Type-specific probes are identified. Also, a type-specific PCR for for example serotype 9 is provided, being a rapid, reliable and sensitive assay, which is used directly on nasal or tonsillar swabs or other samples of infected or carrier animals.

The invention also provides a probe or primer according to the invention further provided with at least one reporter molecule, Examples of reporter molecules are manifold and known in the art, for example a reporter molecule can comprise additional nucleic acid provided with a specific sequence (e.g. oligo-dT) hybridising to a corresponding sequence to which hybridisation can easily be detected for example because it has been immobilised to a solid support.

Yet other reporter molecules comprise chromophores, e.g. fluorochromes for visual detection, for example by light microscopy or fluorescent in situ hybridisation (FISH) techniques, or comprise an enzyme such as horseradish peroxidase for enzymatic detection, e.g in enzyme-linked assays (EIA). Yet other reporter molecules comprise radioactive compounds for detection in radiation-based-assays.

In a preferred embodiment of the invention, at least one probe or primer according to the invention is provided (labelled) with a reporter molecule and a quencher molecule, providing together with unlabeled probe or primer a PCR-based test allowing rapid detection of specific hybridisation.

The invention further provides a diagnostic test or test kit comprising a probe or primer as provided by the invention. Such a test or test kit, for example a cross-hybridisation test or PCR-based test, is advantageously used in rapid detection and/or serotyping of *Streptococcus suis*.

The invention furthermore provides a protein or fragment thereof encoded by a nucleic acid according to the invention. Examples of such a protein or fragment are for example

proteins described in for example Table 2 of the description, for example a cpsA protein is provided functionally encoding regulation of capsular polysaccharide synthesis, whereas cpsB and cpsC are functionally involved in chain in chain length determination. Other proteins or functional fragments thereof as provided by the invention, such as cpsD, E, F, G, H, I, J, K and related proteins, are involved in polysaccharide biosynthesis, functioning for example as glucosyl- or glycosyltransferase in polysaccharide biosynthesis of *Streptococcus suis* capsular antigen.

The invention furthermore provides a method to produce a *Streptococcus suis* capsular antigen comprising using a protein or functional fragment thereof as provided by the invention, and provides therewith a *Streptococcus suis* capsular antigen obtainable by such a method. A comparison of the predicted amino acid sequences of the *cps2* genes with sequences found in the databases allowed the assignment of functions to the open reading frames. The central region contains the type specific glycosyltransferases and the putative polysaccharide polymerase. This region is flanked by two regions encoding for proteins with common functions, such as regulation and transport of polysaccharide across the membrane. Biosynthesis of Streptococcus capsular polysaccharide antigen using a protein or functional fragment thereof is advantageously used in chemo-enzymatic synthesis and the development of vaccines which offer protection against serotype-specific Streptococcal disease, and is also advantageously used in the synthesis and development of multivalent vaccines against Streptococcal infections. Such vaccines elicit anticapsular antibodies which confer protection.

The invention furthermore provides a vaccine comprising an antigen according to the invention and further comprising a suitable carrier or adjuvant. The immunogenicity of a capsular antigen provided by the invention is for example increased by linking to a carrier (such as a carrier protein), allowing the

recruitment of T-cell help in developing an immune response.

The invention further provides a recombinant micro-organism provided with at least a part of a capsular gene cluster derived from *Streptococcus suis*. The invention provides for example a lactic acid bacterium provided with at least a part of a capsular gene cluster derived from *Streptococcus suis*. Various food-grade lactic acid bacteria (Lactococcus lactis, Lactobacillus casei, Lactobacillus plantarium and Streptococcus gordonii) have been used as delivery systems for mucosal immunization. It has now been shown that oral (or mucosal) administration of recombinant L. lactis, Lactobacillus, and Streptococcus gordonii can elicit local IgA and /or IgG antibody responses to an expressed antigen. The use of oral routes for immunization against infective diseases is desirable because oral vaccines are easier to administer, have higher compliance rates, and because mucosal surfaces are the portals of entry for many pathogenic microbial agents. It is within the skill of the artisan to provide such micro-organisms with (additional) genes.

The invention further provides a recombinant *Streptococcus suis* mutant provided with a modified capsular gene cluster. It is within the skill of the artisan to swap genes within a species. In a preferred embodiment, an avirulent *Streptococcus suis* mutant is selected to be provided with at least a part of a modified capsular gene cluster according to the invention.

The invention further provides a vaccine comprising a micro-organism or a mutant provided by the invention. An advantage of such a vaccine over currently used vaccines is that they comprise accurately defined micro-organisms and well-characterised antigens, allowing accurate determination of immune responses against various antigens of choice.

The invention is further explained in the experimental part of this description without limiting the invention thereto.

Experimental part

MATERIAL AND METHODS

Bacterial strains and growth conditions. The bacterial strains and plasmids used in this study are listed in Table 1. *S. suis* strains were grown in Todd-Hewitt broth (code CM189, Oxoid), and plated on Columbia agar blood base (code CM331, Oxoid) containing 6% (v/v) horse blood. *E.coli* strains were grown in Luria broth (28) and plated on Luria broth containing 1.5% (w/v) agar. If required, antibiotics were added to the plates at the following concentrations: spectinomycin: 100 ug/ml for *S. suis* and 50 ug/ml for *E. coli* and ampicillin, 50 ug/ml.

Serotyping. The *S.suis* strains were serotypes by the slide agglutination test with serotype-specific antibodies (44).

DNA techniques. Routine DNA manipulations were performed as described by Sambrook et al. (36).

Alkaline phosphatase activity. To screen for PhoA fusions in *E.coli*, plasmid libraries were constructed. Therefore, chromosomal DNA of *S. suis* type 2 was digested with *Alu*I. The 300-500-bp fragments were ligated to *Sma*I-digested pPHOS2. Ligation mixtures were transformed to the PhoA$^-$ *E. coli* strain CC118. Transformants were plated on LB media supplemented with 5-Bromo-4-chloro-3-indolylfosfaat (BCIP, 50 ug/ml, Boehringer, Mannheim, Germany). Blue colonies were purified on fresh LB/BCIP plates to verify the blue phenotype.

DNA sequence analysis. DNA sequences were determined on a 373A DNA Sequencing System (Applied Biosystems, Warrington, GB). Samples were prepared by use of a ABI/PRISM dye terminator cycle sequencing ready reaction kit (Applied Biosystems). Sequencing data were assembled and analyzed using the MacMollyTetra program. Custom-made sequencing primers were purchased from Life Technologies. Hydrophobic stretches within proteins were predicted by the method of Klein et al. (17). The

BLAST program available on Netscape Navigator™ was used to search for protein sequences related to the deduced amino acid sequences.

Construction of gene-specific knock-out mutants of *S. suis*. To construct the mutant strains 10cpsB and 10cpsEF we electrotransformed the pathogenic serotype 2 strain 10 (45, 49) of *S. suis* with pCPS11 and pCPS28 respectively. In these plasmids the *cps*B and *cps*EF genes were disturbed by the insertion of a spectinomycin-resistance gene. To create pCPS11 the internal 400 bp *Pst*I-*Bam*HI fragment of the *cps*B gene in pCPS7 was replaced by the Spc^R gene. For this purpose pCPS7 was digested with *Pst*I and *Bam*HI and ligated to the 1,200-bp *Pst*I-*Bam*HI fragment, containing the Spc^R gen, from pIC-spc. To construct pCPS28 we have used pIC20R. In this plasmid we inserted the *Kpn*I-*Sal*I fragment from pCPS17 (resulting in pCPS25) and the *Xba*I-*Cla*I fragment from pCPS20 (resulting in pCPS27). pCPS27 was digested with *Pst*I and *Xho*I and ligated to the 1,200-bp *Pst*I-*Xho*I fragment, containing the Spc^R gene of pIC-spc. The electrotransformation to *S. suis* was carried out as described before (38).

Southern blotting and hybridization. Chromosomal DNA was isolated as described by Sambrook *et al.* (36). DNA fragments were separated on 0.8% agarose gels and transferred to Zeta-Probe GT membranes (Bio-Rad) as described by Sambrook *et al.* (36). DNA probes were labelled with [($-^{32}$P]dCTP (3000 Ci $mmol^{-1}$; Amersham) by use of a random primed labelling kit (Boehringer). The DNA on the blots was hybridized at 65°C with appropriate DNA probes as recommended by the supplier of the Zeta-Probe membranes. After hybridization, the membranes were washed twice with a solution of 40 mM sodium phosphate, pH 7.2, 1 mM EDTA , 5% SDS for 30 min at 65°C and twice with a solution of 40 mM sodium phosphate, pH 7.2, 1 mM EDTA, 1% SDS for 30 min at 65°C.

PCR. The primers used in the *cps2J* PCR correspond to the positions 13791-13813 and 14465-14443 in the *S. suis cps2* locus. The sequences were: 5'-CAAACGCAAGGAATTACGGTATC-3' and

5'-GAGTATCTAAAGAATGCCTATTG-3'. The primers used for the *cps1I* PCR correspond to the positions 4398-4417 and 4839-4821 in the *S. suis cps1* sequence. The sequences were: 5'-GGCGGTCTAGCAGATGCTCG-3' and 5'-GCGAACTGTTAGCAATGAC-3'. The primers used in the *cps9H* PCR correspond to the positions 4406-4126 and 4494-4475 in the *S. suis cps9* sequence. The sequences were: 5'-GGCTACATATAATGGAAGCCC3' and 5'-CGGAAGTATCTGGGCTACTG-3'.

Electron Microscopy. Bacteria were prepared for electron microscopy as described by Wagenaar et al. (50). Shortly, bacteria were mixed with agarose MP (Boehringer) of 37° C to a concentration of 0.7%. The mixture was immediately cooled on ice. Upon gelifying, samples were cut into 1 to 1.5 mm slices and incubated in a fixative containing 0.8% glutaraldehyde and 0.8% osmiumtetraoxide. Subsequently, the samples were fixed and stained with uranyl acetate by microwave stimulation, dehydrated and imbedded in eponaraldite resin. Ultra-thin sections were counterstained with lead citrate and examined with a Philips CM 10 electron microscope at 80 kV.

Isolation of porcine alveolar macrophages (AM). Porcine AM were obtained from the lungs of specific pathogen free (SPF) pigs. Lung lavage samples were collected as described by van Leengoed *et al.* (43). Cells were suspended in EMEM containing 6% (v/v) SPF-pig serum and adjusted to 10^7 cells per ml.

RESULTS

Identification of the *cps* locus.

The first part of the *cps* locus of *S.suis* type 2 was identified by making use of a strategy developed for the genetic identification of exported proteins (13, 31). In this system we made use of a plasmid (pPHOS2) containing a truncated alkaline phosphatase gene (13). The gene lacked the promoter sequence, the translational start site and the signal sequence. The truncated gene is precede by a unique *Sma*I restriction site. Chromosomal DNA of *S. suis* type 2, digested with *Alu*I, was

randomly cloned in this restriction site. Because translocation of PhoA across the cytoplasmic membrane of *E. coli* is required for enzymatic activity, the system can be used to select for *S. suis* fragments containing a promoter sequence, a translational start site and a functional signal sequence. Among 560 individual *E. coli* clones tested, 16 displayed a dark blue phenotype when plated on media containing BCIP. DNA sequence analysis of the inserts from several of these plasmids were performed (results not shown) and the deduced amino acid sequences were analyzed. The hydrophobicity profile of one of the clones (pPHOS7, results not shown) showed that the N-terminal part of the sequence resembled the characteristics of a typical signal peptide: a short hydrophilic N-terminal region is followed by a hydrophobic region of 38 amino acids. These data indicate that the phoA system was successfully used for the selection of *S. suis* genes encoding exported proteins. Moreover, the sequences were analyzed for similarities present in the databases. The sequence of pPHOS7 showed a high similarity (37% identity) with the protein encoded by the *cps14C* gene of *Streptococcus pneumoniae* (19). This strongly suggests that pPHOS7 contains a part of the *cps* operon of *S. suis* type 2.

Cloning of the flanking *cps* genes. In order to clone the flanking *cps* genes of *S. suis* type 2 the insert of pPHOS7 was used as a probe to identify chromosomal DNA fragments which contain flanking *cps* genes. A 6-kb *Hind*III fragment was identified and cloned in pKUN19. This yielded clone pCPS6 (Fig. 1C). Sequence analysis of the insert of pCPS6 revealed that pCPS6 most probably contained the 5'-end of the *cps* locus, but still lacked the 3'-end (see below). Therefore, sequences of the 3' -end of pCPS6 were in turn used as a probe to identify chromosomal fragments containing *cps* sequences located further downstream. These fragments were also cloned in pKUN19, resulting in pCPS17. Using the same system of chromosomal walking we subsequently generated the plasmid pCPS18, pCPS20, pCPS23 and pCPS26, containing downstream *cps* sequences.

Analysis of the cps operon. The complete nucleotide sequence of the cloned fragments was determined (figure 4). Examination of the compiled sequence revealed the presence of at least 13 potential open reading frame (Orfs), which were designated as Orf 2Y, Orf2X and Cps2A-Cps2K (Fig. 1A). Moreover, a 14th, incomplete, Orf (Orf 2Z) was located at the 5'-end of the sequence. Two potential promoter sequences were identified. One was located 313 bp (locations 1885-1865 and 1884-1889) upstream of Orf2X. The other potential promoter sequence was located 68 bp upstream of Orf2Y (locations 2241-2236 and 2216-2211). Orf2Y is expressed in opposite orientation. Between Orfs 2Y and 2Z the sequence contained a potential stem-loop structure, which could act as a transcription terminator. Each Orf is preceded by a ribosome-binding site and the majority of the Orfs are very closely linked. The only significant intergenic gap was found between Cps2G and Cps2H (389 nucleotides). However, no obvious promoter sequences or potential stem-loop structures were found in this region. These data suggest that Orf2X and Cps2A-Cps2K are arranged as an operon.

An overview of all Orfs with their properties is shown in Table 2. The majority of the predicted gene products is related to proteins involved in polysaccharide biosynthesis. Orf2Z showed some similarity with the YitS protein of *Bacillus subtilis*. YitS was identified during the sequence analysis of the complete genome of *B. subtilis*. The function of the protein is unknown.

Orf2Y showed similarity with YcxD protein of *B. subtilis* (53). Based on the similarity between YcxD and MocR of *Rhizobium meliloti* (33), YcxD was suggested to be a regulatory protein.

Orf2X showed similarity with the hypothetical YAAA proteins of *Haemophilus influenzae* and *E. coli*. The function of these proteins is unknown.

The gene products encoded by the *cps2A, cps2B, cps2C* and *cps2D* genes showed approximate similarity with the CpsA, CpsC,

CpsD and CpsB proteins of several serotypes of *Streptococcus pneumoniae* (19),, respectively. This suggest similar functions for these proteins. Hence, Cps2A may have a role in the regulation of the capsular polysaccharide synthesis. Cps2B and Cps2C could be involved in the chain length determination of the type 2 capsule and Cps2C can play an additional role in the export of the polysaccharide. The Cps2D protein of *S. suis* is related to the CpsB protein of *S. pneumoniae* and to proteins encoded by genes of several other Gram-positive bacteria involved in polysaccharide or exopolysaccharide synthesis, but their function is unknown (19).

The protein encoded by *cps2E* gene showed similarity to several bacterial proteins with glycosyl transferase activities: Cps14E and Cps19fE of *S. pneumoniae* serotypes 14 and 19F (18, 19, 29), CpsE of *Streptococcus salvarius* (X94980) and CpsD of *Streptococcus agalactiae* (34). Recently, Kolkman *et al.* (18) showed that Cps14E is a glucosyl-1-phosphate transferase that links glucose to a lipid carrier, the first step in the biosynthesis of the S. *pneumoniae* type 14 repeating unit. Based on these data a similar function may be fulfilled by Cps2E of *S. suis* .

The protein encoded by the *cps2F* gene showed similarity to the protein encoded by the *rfbU* gene of *Salmonella enteritica.* (25). This similarity is most pronounced in the C-terminal regions of these proteins. The *rfbU* gene was shown to encoded mannosyltransferase activity (25).

The *cps2G* gene encoded a protein that showed moderate similarity with the *rfbF* gene product of *Campylobacter hyoilei* (22), the *eps*F gene product of *S. thermophilus* (40) and the *capM* gene product of *S. aureus* (24). On the basis of similarity the *rfbF, epsF* and *capM* genes are suggested to encoded galactosyltransferase activities. Hence, a similar glycosyl transferase activity could be fulfilled by the *cps2G* gene product.

The *cps2H* gene encodes a protein that is similar to the N-terminal region of the *lgtD* gene product of *Haemophilus*

influenzae (U32768). Moreover, the hydrophobicity plots of Cps2H and LgtD looked very similar in these regions (data not shown). Based on sequence similarity the *lgtD* gene product was suggested to have glycosyl transferase activity (U32768).

The gene product encoded by the *cps2I* gene showed some similarity with a protein of *Actinobacillus actinomycetemcomitans* (AB002668). This protein is part of the gene cluster responsible for the serotype-b-specific antigen of *A. actimycetemcomitans*. The function of the protein is unknown.

The gene products encoded by the *cps2J* and *cps2K* genes showed significant similarities to the Cps14J protein of *S. pneumoniae*. The *cps14J* gene of *S. pneumoniae* was shown to encode a ß-1,4-galactosyltransferase activity. In *S. pneumoniae* CpsJ is responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide (20). Even some similarity was found between Cps2J and Cps2K (Fig. 2, 25.5% similarity). This similarity was most pronounced in the N-terminal regions of the proteins. Recently, two small conserved regions were identified in the N-terminus of Cps14J and Cps14I and their homologues (20). These regions were predicted to be important for catalytic activity. Both regions, DXS and DXDD (Fig. 2), were also found in Cps2J and Cps2K.

Distribution of the *cps2* genes in other *S. suis* serotypes. To examine the relationship between the *cps2* genes and *cps* genes in the other *S. suis* serotypes, we performed cross-hybridization experiments. DNA fragments of the individual *cps2* genes were amplified by PCR, labelled with ^{32}P, and used to probe Southern blots of chromosomal DNA of the reference strains of the 35 different *S. suis* serotypes. Large variation in the hybridization patterns were observed (Table 4). As a positive control we used a probe specific for 16S rRNA. The 16S rRNA probe hybridized with all serotypes tested. However, none of the other genes tested were common in all serotypes. Based on the genetic organization of the genes we previously

suggested that *orfX* and *cpsA-cpsK* genes are part of one operon and that the protein encoded by these genes are all involved in polysaccharide biosynthesis. OrfY and OrfZ are not a part of this operon, and their role in the polysaccharide biosynthesis is unclear. Based on sequence similarity data, OrfY may be involved in regulation of the *cps2* genes. OrfZ is proposed to be unrelated to polysaccharide biosynthesis. Probes specific for the *orfZ, orfY, orfX, cpsA, cpsB, cpsC* and *cpsD* genes hybridized with most other serotypes. This suggests that the protein encoded by these genes are not type-specific, but may perform more common functions in biosynthesis of the capsular polysaccharide. This confirms previous data which showed that the *cps2A-cps2D* genes showed strong similarity to *cps* genes of several serotype of *Streptococcus pneumoniae*. Based on this similarity Cps2A is possibly a regulatory protein, whereas Cps2B and Cps2C may play a role in length determination and export of polysaccharide. The *cps2E* gene hybridized with DNA of serotypes 1, 2, 14 and 1/2. The *cps2E* gene showed a strong similarity to the *cps14E* gene of *S. pneumoniae* (18). This enzyme was shown to have a glucosyl-1-phosphate activity and catalyzed the transfer of glucose to a lipid carrier (18). These data indicate that a glycosyltransferase closely related to Cps14E may be responsible for the first step in the biosynthesis of polysaccharide in the *S. suis* serotypes 1, 2, 14 and 1/2. The *cps2F, cps2G, cps2H, cps2I* and *cps2J* genes hybridized with chromosomal DNA of serotypes 2 and 1/2 only. The *cps2G* gene showed an additional weak hybridization signal with DNA of serotype 34. In agglutination tests serotype 1/2 showed agglutination with sera specific for serotype 2 as well as with sera specific for serotype 1. This suggests that serotype 1/2 shares antigenic determinants with both types 1 and 2. The hybridization data confirmed these data. All putative glycosyltransferases present in serotype 2 are also present in serotype 1/2. The *cps2K* gene showed a similar hybridization pattern as the *cps2E* gene. Hybridization was observed with DNA

of serotypes 1, 2, 14 and 1/2. Taken together these hybridization data show that the *cps2* gene cluster can be divided in three regions: a central region containing the type-specific genes is flanked by two regions containing common genes for various serotypes.

Cloning of the type-specific *cps* genes of serotypes 1 and 9. To clone the type-specific *cps* genes of *S. suis* serotype 1 we used the *cps2E* gene as a probe to identify chromosomal DNA fragments of type 1 which contain flanking *cps* genes. A 5 kb *Eco*RV fragment was identified and cloned in pKUN19. This yielded pCPS1-1 (Fig. 1B). This fragment was in turn used as a probe to identify an overlapping 2.2 kb *Hin*dIII fragment. pKUN19 containing this *Hin*dIII fragment was designated pCPS1-2. The same strategy was followed to identify and clone the type-specific *cps* genes of serotype 9. In this case, we used the *cps2D* gene as a probe. A 0.8 kb *Hin*dIII-*Xba*I fragment was identified and cloned, yielding pCPS9-1 (Fig. 1C). This fragment was in turn used as a probe to identify a 4 kb *Xba*I fragment. pKUN19 containing this 4 kb *Xba*I fragment was designated pCPS9-2.

Analysis of the cloned *cps1* genes. The complete nucleotide sequence of the inserts of pCPS1-1 and pCPS1-2 was determined (figure 5). Examination of the sequence revealed the presence of five complete and two incomplete Orfs (Fig.1B). Each Orf is preceded by a ribosome-binding site. In accord with data obtained for the *cps2* genes of serotype 2, the majority of the Orfs is very closely linked. The only significant gap (718 bp) was found between Cps1G and Cps1H. No obvious promoter sequences or potential stem-loop structures could be found in this region. This suggests that, as in serotype 2, the *cps* genes in serotype 1 are arranged in an operon.

An overview of the Orfs and their properties in shown in Table 2. As expected on the basis of the hybridization data

(Table 4), the protein encoded by the *cps1E* gene was related to Cps2E of *S. suis* type 2 (identity of 86%). The fragment cloned in pCPS1-1 lacked the coding region for the first 7 amino acids of the *cps1E* gene.

The protein encoded by the *cps1F* and *cps1G* genes showed strong similarity to the Cps14F and Cps14G proteins of *Streptococcus pneumoniae* serotype 14, respectively (20). The function of the Cps14F is not completely clear, but it has been suggested that Cps14F can enhance role in glycosyltransferase activity. The *cps14G* gene of *S. pneumoniae* was shown to encode ß-1,4-galactosyltransferase activity. In *S. pneumoniae* type 14 this activity is required for the second step in the biosynthesis of the oligosaccharide subunit (20). Based on the similarity data found similar glycosyltransferase and enhancing activities are suggested for the *cps 1G* and *cps1F* genes of *S. suis* type 1.

The protein encoded by the *cps1H* gene showed similarity to the Cps14H protein of *S. pneumoniae* (20). Based on sequence similarity Cps14H was proposed to be the polysaccharide polymerase (20).

The protein encoded by the *cps1I* gene showed some similarity with the Cps14J protein of *S. pneumoniae* (19). The *cps14J* gene was shown to encode a ß-1,4-galactosyltransferase activity, responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide.

Between Cps1G and Cps1H a gap of 718 bp was found. This region revealed three small Orfs. The three Orfs were expressed in three different reading frames and were not preceded by potential ribosome binding sites, nor contained potential start sites. However, the three potential gene products encoded by this region showed some similarity with three successive regions of the C-terminal part of the EpsK protein of *Streptococcus thermophilus* (27% identity, 40). The region related to the first 82 amino acids is lacking.

Analysis of the cloned *cps9* genes. We also determined the complete nucleotide sequence of the inserts of pCPS9-1 and pCPS9-2 (figure 6). Examination of the sequence revealed the presence of three complete and two incomplete Orfs (Fig. 1C). As in serotypes 1 and 2, all Orfs are preceded by a ribosome-binding site and are very closely coupled. As suggested by the hybridization data (Table 4) the Cps2D and Cps9D proteins were highly related (Table 2). Based on sequence comparisons pCPS9-1 lacked the first 27 amino acids of the Cps9D protein.

The protein encoded by the *cps9E* gene showed some similarity with the CapD protein of *Staphylococcus aureus* serotype 1 (24). Based on sequence similarity data the Cap1D protein was suggested to be an epimerase or a dehydratase involved in the synthesis of N-acetylfructosamine or N-acetylgalactosamine (63).

Cps9F showed some similarity to the CapM proteins of *S. aureus* serotypes 5 and 8 (61, 64, 65). Based on sequence similarity data Cap5M and Cap8M are proposed to be glycosyltransferases (63).

The protein encoded by the *cps9G* gene showed some similarity with a protein of *Actinobacillus actinomycetemcomitans* (AB002668_4). This protein is part of a gene cluster responsible for the serotype-b specific antigens of *Actinobacillus actinomycetemcomitans* . The function of the protein is unknown.

The protein encoded by the *cps9H* gene showed some similarity with the *rfbB* gene of *Yersinia enterolitica* (68). The RfbB protein was shown to be essential for O-antigen synthesis, but the function of the protein in the synthesis of the O:3 lipopolysaccharide is unknown.

Serotype 1 and serotype 9 specific *cps* genes. To determine whether the cloned fragments in pCPS1-1, pCPS1-2, pCPS9-1 and pCPS9-2 contained the type-specific genes for serotype 1 and 9, respectively, cross hybridization experiments were

performed. DNA fragments of the individual *cps1* and *cps9* genes were amplified by PCR, labelled with ^{32}P, and used to probe Southern blots of chromosomal DNA of the reference strains of the 35 different *S. suis* serotypes. The results are shown in Table 5. Based on the data obtained with the *cps2E* probe (Table 4), the *cps1E* probe was expected to hybridize with chromosomal DNA of *S. suis* serotypes 1,2, 14, 27 and 1/2. The *cps1H, cps9E* and *cps9F* probes hybridized with most other serotypes. However, the *cps1F* and *cps1G* and cps1I probes hybridized with chromosomal DNA of serotypes 1 and 14 only. The *cps9G* and *cps9H* probe hybridized with serotype 9 only. These data suggest that the *cps9G* and cps9H probes are specific for serotype 9 and therefore could be useful tools for the development of rapid and sensitive diagnostic tests for *S. suis* type 9 infections.

Type specific PCR. So far, the probes were tested on the 35 different reference strains only. To test the diagnostic value of the type-specific *cps* probes further, several other *S. suis* serotype 1, 2, 1/2, 9 and 14 strains were used. Moreover, since a PCR based method would be even more rapid and sensitive than a hybridization test, we tested whether we could use a PCR for the serotyping of the *S. suis* strains. The oligonucleotide primer sets were chosen within the *cps2J, cps1I* and *cps9H* genes. Amplified fragments of 675 bp, 380 bp and 390 bp were expected respectively. The results show that 675 bp fragments were amplified on type 2 and 1/2 strains using *cps2J* primers; 380 bp fragments were amplified on type 1 and 14 strains using *cps1I* primers and 390 bp fragments were amplified on type 9 strains using *cps9H* primers.

DISCUSSION

We describe the identification and the molecular characterisation of the *cps* locus, involved in the capsular

polysaccharide biosynthesis, of *S. suis* serotype 2. A region of 16 kb was cloned and sequenced. 14 open reading frames were identified. Most of the genes seemed to belong to a single transcriptional unit, suggesting a co-ordinate control of these genes. We assign functions to most of the gene products. We thereby identified regions involved in regulation (Cps2A), chain length determination (Cps2B, C), export (Cps2C) and biosynthesis (Cps2E, F, G, H, J, K). The region involved in biosynthesis is located at the centre of the gene cluster and is flanked by two regions containing genes with more common functions. The incomplete *orf2Z* gene was located at the 5'-end of the cloned fragment. Orf2Z showed some similarity with the YitS protein of *B.subtilis*. However, because the function of the YitS protein is unknown this did not give us any information about the possible function of Orf2Z. Because the *orf2Z* gene is not a part of the *cps* operon, a role of this gene in polysaccharide biosynthesis is not expected. The Orf2Y protein showed some similarity with the YcxD protein of *B.subtilis* (53).The YcxD protein was suggested to be a regulatory protein. Similarly, Orf2Y may be involved in the regulation of polysaccharide biosynthesis. The Orf2X protein showed similarity with the YAAA proteins of *H. influenzae* and *E. coli*. The function of these proteins is unknown. In *S. suis* type 2 the *orf2X* gene seemed to be the first gene in the *cps2* operon. This suggests a role of Orf2X in the polysaccharide biosynthesis. In *H. influenzae* and *E. coli*, however, these proteins are not associated with capsular gene clusters. The analysis of isogenic mutants impaired in the expression of Orf2X should give more insight in the presumed role of Orf2X in the polysaccharide biosynthesis of *S. suis* type 2.

The gene products encoded by the *cps2E, cps2F, cps2G, cps2H, cps2J* and *cps2K* genes showed little similarity with glycosyltransferases of several Gram-positive or Gram-negative bacteria (18, 19, 20, 22, 25). The *cps2E* gene product shows some similarity with the Cps14E protein of *S. pneumoniae* (18, 19). Cps14E is a glucosyl-1-phosphate transferase that links

glucose to a lipid carrier (18). In *S. pneumoniae* this is the first step in the biosynthesis of the oligosaccharide repeating unit. The structure of the *S. suis* serotype 2 capsule contains glucose, galactose, rhamnose, N-acetyl glucoseamine and sialic acid in a ratio of 3:1:1:1:1 (7). Based on these data we conclude that Cps2E of *S. suis* has glucosyltransferase activity, and is involved in the linkage of the first sugar to the lipid carrier.

The C-terminal region of the *cps2F* gene product showed some similarity with the RfbU of *Salmonella enteritica*. RfbU was shown to have mannosyltransferase activity (24). Because mannosyl is not a component of the *S. suis* type 2 polysaccharide a mannosyltransferase activity is not expected in this organism. Nevertheless, *cps2F* encodes a glycosyltransferase with another sugar specificity.

Cps2G showed moderate similarity to a family of gene products suggested to encode galactosyltransferase activities (22, 24, 40). Hence a similar activity is shown for Cps2G.

Cps2H showed some similarity with LgtD of *H. influenzae* (U32768). Because LgtD was proposed to have glycosyltransferase activity , a similar activity is fulfilled by Cps2H.

Cps2J and Cps2K showed similarity to Cps14J of *S. pneumoniae* (20). Cps2J showed similarity with Cps14I of *S. pneumoniae* as well. Cps14I was shown to have N-acetyl glucosaminyltransferase activity, whereas Cps14J has a ß-1,4-galactosyltransferase activity (20). In *S. pneumoniae* Cps14I is responsible for the addition of the third sugar and Cps14J for the addition of the last sugar in the synthesis of the type 14 repeating unit (20). Because the capsule of *S. suis* type 2 contains galactose as well as N-acetyl glucosamine components, galactosyltransferase as well as N-acetyl glucoaminyltransferase activities could be envisaged for the *cps2J* and *cps2K* gene products, respectively. As was observed for Cps14I and Cps14J, the N-termini of Cps2J and Cps2K showed a significant degree of sequence similarity. Within the N-terminal domains of Cps14I and Cps14J, two small regions were

identified, which were also conserved in several other glycosyltransferases (22). Within these two regions, two Asp residues were proposed to be important for catalytic activity. The two conserved regions, DXS and DXDD, were also found in Cps2J and Cps2K.

The function of Cps2I remains unclear. Cps2I showed some similarity with a protein of *A. actinomycetemcomitans*. Although this protein part is of the gene cluster responsible for the serotype-B-specific antigens, the function of the protein is unknown.

We further describe the identification and characterization of the *cps* genes specific for *S. suis* serotypes 1, 2 and 9. After the entire *cps2* locus of *S. suis* serotype 2 was cloned and characterized, functions for most of the *cps2* gene products could be assigned by sequence homologies. Based on these data the glycosyltransferase activities, required for type specificity, could be located in the centre of the operon. Cross-hybridization experiments, using the individual *cps2* genes as probes on chromosomal DNAs of the 35 different serotypes, confirmed this idea. The regions containing the type-specific genes of serotypes 1 and 9 could be cloned and characterized, showing that an identical genetic organization of the *cps* operons of other *S. suis* serotypes exists. The *cps1E, cps1F, cps1G, cps1H,* and *cps1I* genes revealed a striking similarity with *cps14 E, cps14F, cps14G, cps14H* and *cps14J* genes of *S. pneumoniae*. Interestingly, *S. pneumoniae* serotype 14 is the serotype most commonly associated with pneumococcal infections in young children (54), whereas *S. suis* serotype 1 strains are most commonly isolated from piglets younger than 8 weeks (46). In *S. pneumoniae* the *cps14E, cps14G, cps14I* and *cps14J* encode the glycosyltransferases required for the synthesis of the type 14 tetrameric repeating unit, showing that the *cps1E, cps1G* and *cps1I* genes encoded glycosyltransferases. The precise functions of these genes as well as the substrate

specificities of the enzymes can be established. In *S. pneumoniae* the *cps14E* gene was shown to encode a glucosyl-1-phosphate transferase catalyzing the transfer of glucose to a lipid carrier. Moreover, *cpsE*-like genes were found in *S. pneumoniae* serotypes 9N, 13, 14, 15B, 15C, 18F, 18A and 19F (60). *CpsE* mutants were constructed in the serotypes 9N, 13 , 14 and 15B. All mutant strains lacked glucosyltransferase activity (60). Moreover, in all these *S. pneumoniae* serotypes the *cpsE* gene seemed to be responsible for the addition of glucose to the lipid carrier. Based on these data we suggest that in *S. suis* type 1 the *cps1E* gene may fulfil a similar function. The structure of the *S. suis* type 1 capsule is unknown, but it is composed of glucose, galactose, N-acetyl glucosamine, N-acetyl galactosamine and sialic acid in a ratio of 1: 2.4: 1: 1:1.4 (5). Therefore a role of a *cpsE*-like glucosyltransferase activity can easily be envisaged. *CpsE* like sequences were also found in serotypes 2, 1/2 and 14.

For polysaccharide biosynthesis in *S. pneumoniae* type 14, transfer of the second sugar of the repeating unit to the first lipid-linked sugar is performed by the gene products of *cps14F* and *cps14G* (20). Similar to Cps14F and Cps14G, the *S. suis* type 1 proteins Cps1F and Cps1G may act as one glycosyltransferase performing the same reaction. Cps14F and Cps14G of *S. pneumoniae* showed similarity to the N-terminal half and C-terminal half of the SpsK protein of *Sphingomonas* (20, 67), respectively. This suggests a combined function for both proteins. Moreover, *cps14F* and *cps14G* like sequences were found in several serotypes of *S. pneumoniae* and these genes always seemed to exist together (60). The same was observed for *S. suis type* 1. The *cps1F* and *cps1G* probes hybridized with type 1 and type 14 strains.

According to the similarity found between the *cps1H* gene and the *cps14H* gene of *S. pneumoniae* (20), *cps1H* is expected to encode a polysaccharide polymerase.

The protein encoded by the *cps1I* gene showed some similarity with the Cps14J protein of *S. pneumoniae* (19). The

cps14J gene was shown to encode a ß-1,4-galactosyltransferase activity, responsible for the addition of the fourth (i.e. last) sugar in the synthesis of the *S. pneumoniae* serotype 14 polysaccharide. In *S. suis* type 2 the proteins encoded by the *cps2J* and *cps2K* genes showed similarity to the Cps14J protein. However, no significant homologies were found between Cps2J, Cps2K and Cps1I. In the N-terminal regions of Cps14J and Cps14I two small conserved regions, DXS and DXDD, were identified (19). These regions seemed to be important for catalytic activity (13). At the same positions in the sequence Cps2I contained the regions DXS and DXED.

In the region between Cps1G and Cps1H three small Orfs were identified. Since the Orfs were expressed in three different reading frames, and did not contain potential start sites, expression is not expected. However, the three potential gene products encoded by this region showed some similarity with three successive regions of the C-terminal part of the EpsK protein of *Streptococcus thermophilus* (27% identity, 40). The region related to the first 82 amino acids is lacking. The EpsK protein was suggested to play a role in the export of the exopolysaccharide by rendering the polymerized exopolysaccharide more hydrophobic through a lipid modification. These data could suggest that the sequences in the region between Cps1G and Cps1H originated from *epsK*-like sequence. Hybridization experiments showed that this *epsK*-like region is also present in other serotype 1 strains as well as in serotype 14 strains (results not shown).

The function of most of the cloned serotype 9 genes can be established. Based on sequence similarity data the *cps9E* and *cps9F* genes could be glycosyltransferases (61, 24, 63, 64, 65). Moreover, the *cps9G* and *cps9H* genes showed similarity to genes located in regions involved in polysaccharide biosynthesis, but the function of these genes is unknown (68).

Cross-hybridization experiments using the individual *cps2*, *cps1* and *cps9* genes as probes showed that the *cps9G* and *cps9H* probes specifically hybridized with serotype 9 strains.

Therefore, these are useful as tools for the identification of *S. suis* type 9 strains both for diagnostic purposes as well as in epidemiological and transmission studies. We previously developed a PCR method which can be used to detect *S. suis* strains in nasal and tonsil swabs of pigs (62). The method was for example used to identify pathogenic (EF-positive) strains of *S. suis* serotype 2 During the last years, beside *S. suis* type 2 strains, serotype 9 strains are frequently isolated from organs of diseased pigs. However, until now a rapid and sensitive diagnostic test was not available for type 9 strains. Therefore, the type 9 specific probes or the type 9 specific PCR is of great diagnostic value. The *cps1F, cps1G* and *cps1I* probes hybridized with serotype 1 as well as with serotype 14 strains. In coagglutination tests type 1 strains react with the anti-type 1 as well as with the anti-type 14 antisera (56). This suggests the presence of common epitopes between these serotypes. On the other hand type 1 strains agglutinated only with anti-type 1 serum (56,57), indicating that it is possible to detect differences between those serotypes.

The *cps2F, cps2G, cps2H, cps2I* and *cps2J* probes hybridized with serotypes 2 and 1/2 only. Serotype 34 showed a weak hybridizing signal with the *cps2G* probe. As shown in agglutination tests type 1/2 strains react with sera directed against type 1 as well as with sera directed against type 2 strains (46). Therefore, type 1/2 shared antigens with both types 1 and 2. Based on the hybridization patterns of serotype 1/2 strains with the *cps1* and *cps2* specific genes, serotype 1/2 seemed to be more closely related to type 2 strains than to type 1 strains. In our current studies we identify type-specific genes, primers or probes which are used for the discrimination of serotypes 1, 14 and 2 and 1/2 and others of the 35 serotypes yet known. Furthermore, type-specific genes, primers or probes can now easily be developed for yet unknown serotypes, once they become isolated.

TABLE 1. **Bacterial strains and plasmids**

strain/plasmid	source/reference	relevant characteristics
Strain		
E.coli		
CC118	PhoA⁻	(28)
XL2 blue	Stratagene	
E.coli		
XL2 blue	Stratagene	
S. suis		
10	virulent serotype 2 strain	(49)
3	serotype 2	(63)
17	serotype 2	(63)
735	reference strain serotype 2	(63)
T15	serotype 2	(63)
6555	reference strain serotype 1	(63)
6388	serotype 1	(63)
6290	serotype 1	(63)
5637	serotype 1	(63)
5673	serotype 1/2	(63)
5679	serotype 1/2	(63)
5928	serotype 1/2	(63)
5934	serotype 1/2	(63)
5209	reference strains serotype 1/2	(63)
5218	reference strain serotype 9	(63)
5973	serotype 9	(63)
6437	serotype 9	(63)
6207	serotype 9	(63)
reference strains	serotypes 1-34	(9, 56, 14)
S. suis		
10	virulent serotype 2 strain	(51)
10cpsB	isogenic cpsB mutant of strain 10	this work
10cpsEF	isogenic cpsEF mutant of strain 10	this work
Plasmid		
pKUN19	replication functions pUC, Amp^R	(23)
pGEM7Zf(+)	replication functions pUC, Amp^R	Promega Corp.
pIC19R	replication functions pUC, Amp^R	(29)
pIC20R	replication functions pUC, Amp^R	(29)
pIC-spc	pIC19R containing spc^R gene of pDL282	labcollection
pDL282	replication functions of pBR322 and pVT736-1, Amp^R, Spc^R	(43)
pPHOS2	pIC-spc containing the truncated *pho*A gene of pPHO7 as a *Pst*I-*Bam*HI fragment	this work
pPHO7	contains truncated *pho*A gene	(15)
pPHOS7	pPHOS2 containing chromosomal *S. suis* DNA	this work
pCPS6	pKUN19 containing 6 kb *Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS7	pKUN19 containing 3,5 kb *Eco*RI-*Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS11	pCPS7 in which 0.4 kb *Pst*I-*Bam*HI fragment of *cps*B gene is replaced by Spc^R gene of pIC-spc	this work (Fig.1)
pCPS17	pKUN19 containing 3.1 kb *Kpn*I fragment of *cps* operon	this work (Fig.1)
pCPS18	pKUN19 containing 1.8 kb *Sna*BI fragment of *cps* operon	this work (Fig.1)
pCPS20	pKUN19 containing 3.3 kb *Xba*I-*Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS23	pGEM7Zf(+) containing 1.5 kb *Mlu*I fragment of *cps* operon	this work (Fig.1)
pCPS25	pIC20R containing 2.5 kb *Kpn*I-*Sal*I fragment of pCPS17	this work (Fig.1)
pCPS26	pKUN19 containing 3.0 kb *Hin*dIII fragment of *cps* operon	this work (Fig.1)
pCPS27	pCPS25 containing 2.3 kb *Xba*I (blunt)-*Cla*I fragment of pCPS20	this work (Fig.1)

pCPS28	pCPS27 containing the 1.2 kb *Pst*I-*Xho*I Spc^R gene of pIC-spc	this work (Fig.1)
pCPS29	pKUN19 containing 2.2 kb *Sac*I-*Pst*I fragment of cps operon	this work (Fig.1)
pCPS1-1	pKUN19 containing 5 kb *EcoRV* fragment of *cps* operon of type 1	this work (Fig.1)
pCPS1-2	pKUN19 containing 2.2 kb *Hind*III fragment of *cps* operon of type 1	this work (Fig.1)
pCPS9-1	pKUN19 containing 1 kb *Hind*III-*Xba*I fragment of *cps* operon of serotype 9	this work (Fig.1)
pCPS9-2	pKUN19 containing 4.0 kb *Xba*I-*Xba*I fragment of *cps* operon of serotype 9	this work (Fig.1)

Amp^R: ampicillin resistant
Spc^R: spectinomycin resistant
cps: capsular polysaccharide

TABLE 2. Properties of ORFs in the *cps* locus of *S. suis* serotype 2 and silimarities to gene products of other bacteria

ORF	nucleotide position in sequence	G + C%	number of amino acids	predicted mol. mass (kDa)	predicted pI	proposed function of gene product[1]	imilar gene product (% identity)	reference
ORF Z	? -719		?				*Bacillus subtilis* YitS (26%)	(Y09478)
ORF Y	2079-822	37.9	419	49.4	8.0		*Bacillus subtilis YcxD* (39%)	(53)
ORF X	2202-2934	38.5	244	28.4	8.1		*Haemophilus influenzae* YAAA (24%)	(P43908)
							Escherichia coli YAAA (21%)	(P11288)
Cps2A	3041-4484	38.7	481	53.3	7.9	Regulation	*Streptococcus pneumoniae* Cps19fA (58%)	(12, 29)
							Streptococcus pneumoniae Cps14A (57%)	(19)
							Streptococcus pneumoniae Cap1A (57%)	(30)
							Streptococcus thermophilus EpsA (50%)	(40)
							Streptococcus salvarius CpsA,[c] (56%)	(X94980)
Cps2B	4504-5191	40.1	229	25.2	7.6	Chain length	*Streptococcus pneumoniae* type 3 Orf1 (58%)	(2)
						determination	*Streptococcus pneumoniae* Cap1C (58%)	(30)
							Streptococcus pneumoniae Cps14C (58%)	(19)
							Streptococcus pneumoniae Cps19fC (58%)	(12, 29)
							Streptococcus thermophilus EpsC (54%)	(40)
							Streptococcus salvarius CpsC (54%)	(X94980)
							Streptococcus agalactiae CpsB (44%)	(34)
Cps2C	5203-5878	40.2	225	24.4	8.0	Chain length determination/ Export	*Streptococcus pneumoniae* Cps19fD (60%)	(12, 29)
							Streptococcus pneumoniae Cps14D (59%)	(19)
							Streptococcus pneumoniae Cap1D (60%)	(30)
							Streptococcus agalactiae CpsC (53%)	(34)
							Streptococcus salvarius CpsD (52%)	(X94980)
							Streptoccocus thermophilus EpsD (51%)	(40)
							Lactococcus lactis EpsB (37%)	(42)
Cps2D	5919-6648	38.0	243	28.2	8.0	Unknown	*Streptococcus pneumoniae* Cps19fB (59%)	(12, 29)
							Streptococcus agalactiae CpsA (58%)	(34)
							Streptococcus salvarius CpsB (58%)	(X94980)
							Streptococcus thermophilus EpsB (58%)	(40)
							Streptococcus pneumoniae Cps14B (57%)	(19)
Cps2E	6675-8052	33.4	459	52.9	8.0	Glucosyltransferase	*Streptococcus pneumoniae* Cps14E (56%)	(18, 19)

							Streptococcus salvarius CpsE (56%)	(X94980)
							Streptococcus pneumoniae Cps19fE (55%)	(29)
							Streptococcus agalactiae CpsD (48%)	(34)
Cps2F	8089-9256	32.4	389	45.5	7.8	Glycosyltransferase	*Salmonella enteritica RfbU* (25%)	(25)
Cps2G	9262-10417	35.9	385	43.6	7.9	Glycosyltransferase	*Campylobacter hyoilei* RfbF (25%)	(22)
							Streptococcus thermophilus EpsF (25%)	(40)
							Staphylococcus aureus Cap1M,[C] (25%)	(24)
							Streptococcus thermophilus EpsG (23%)	(40)
Cps2H	10808-12176	31.0	457	53.3	7.9	Glycosyltransferase	*Haemophilus influenzae* LgtD,[N] (28%)	(U32768)
Cps2I	12213-13443	28.8	410	46.9	8.9	Glycosyltransferase	*Actinobacillus actinomycetemcomitans* (28%)	(AB002668)
Cps2J	13583-14579	28.9	332	38.8	7.7	Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (31%)	(20)
							Streptococcus pneumoniae Cps14I (27%)	(20)
							Streptococcus thermophilus EpsI (29%)	(40)
							Lactococcus lactis EpsG,[N] (39%)	(42)
Cps2K	14574-?		?			Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (44%)	(20)
							Streptococcus thermophilus EpsI (39%)	(40)
							Lactococcus lactis EpsG (39%)	(42)

[1]Predicted by sequence similarity
[N] Similarity refers to the amino-terminal part of the gene product
[C] Similarity refers to the carboxy-terminal part of the gene product

TABLE 3. Properties of ORFs in the *cps* genes of *S. suis* serotypes 1 and 9 and silimarities to gene products of other bacteria

ORF	nucleotide position in sequence	G + C%	number of amino acids	predicted mol. mass (kDa)	predicted pI	proposed function of gene product[1]	similar gene product (% identity)	reference/ accession nr.
Cps1E[2]	1-1363	34%	454	52.2	8.0	Glucosyltransferase	*Streptococcus suis* Cps2E (86%)	(26)
(48%)							*Streptococcus pneumoniae* Cps14E	(12)
Cps1F	1374-1821	33%	149	17.3	8.2	Unknown	*Streptococcus pneumoniae* Cps14F (83%)	(14)
Cps1G	1823-2315	25%	164	19.5	7.5	Glycosyltransferase	*Streptococcus pneumoniae* Cps14G(50%)	(14)
Cps1H	3035-4202	24%	389	45.5	8.4	CP polymerase	*Streptococcus pneumoniae* Cps14H (30%)	(14)
Cps1I	4197-					Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (38%)	(13)
							Lactococcus lactis EpsG (31%)	(29)
							Streptococcus thermophilus EpsI (33%)	(28)
Cps1J						Glycosyltransferase	*Streptococcus pneumoniae* Cps14J ( ) (13)	
Cps1K[3]		37%	278	32.5	7.8	Glycosyltransferase	*Streptococcus pneumoniae* Cps14J (44%)	(13)
Cps9D[2]	1-646	37%	215	24.9	8.1	Unknown	*Streptococcus suis* Cps2D (89%)	(26)
Cps9E	680-					Glycosyltransferase	*Staphylococcus aureus* Cap1D (27%)	(18)
Cps9F		36%	200	22.3	8.2	Glycosyltransferase	*Staphylococcus aureus* Cap5M (52%)	(17)
Cps9G		35%	269	31.5	8.0	Unknown	*Actinobacillus actinomycetemcomitans*	

								(43%)	(AB002668_4)
								Haemophilus influenzae Lsg (43%)	(O05081)
Cps9H[3]	30%	143	16.5	7.2	Unknown	. .	. .	*Yersinia enterolitica* RfbB (28%)	(33)

[1]Predicted by sequence similarity
[2] N-terminal part of protein is lacking
[3] C-terminal part of protein is lacking

Table 4. Hybridization of serotype 2*cps* genes and neighbouring sequences with chromosomal DNA of other *S. suis* serotypes

Serotype	DNA probes														
	Z	Y	X	*cpsA2*	*cps2B*	*cpsC2*	*cps2D*	*cps2E*	*cps2F*	*cps2G*	*cps2H*	*cps2I*	*cps2J*	*cps2K*	16rRNA
1	+	+	+	+	+	+	+	+	-	-	-	-	-	+	+
2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
4	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
5	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
6	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
7	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
8	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
9	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
10	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
11	+	+	+	+	-	+	+	-	-	-	-	-	-	-	+
12	+	+	+	+	-	-	±	-	-	-	-	-	-	-	+
13	±	±	±	+	±	±	±	-	-	-	-	-	-	-	+
14	+	+	+	+	+	+	+	+	-	-	-	-	-	+	+
15	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+
16	+	+	+	+	-	±	±	-	-	-	-	-	-	-	+
17	+	+	+	+	±	-	+	-	-	-	-	-	-	-	+
18	+	+	+	+	±	-	+	-	-	-	-	-	-	-	+
19	+	+	+	+	±	-	+	-	-	-	-	-	-	-	+
20	-	±	-	-	-	-	-	-	-	-	-	-	-	-	+
21	+	+	±	+	±	-	±	-	-	-	-	-	-	-	+
22	-	±	-	-	-	-	-	-	-	-	-	-	-	-	+
23	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
24	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
25	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
26	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+
27	+	+	+	+	-	+	+	+	-	-	-	-	-	-	+
28	+	+	+	+	-	±	+	-	-	-	-	-	-	-	+
29	+	+	+	+	+	-	+	-	-	-	-	-	-	-	+
30	+	+	+	+	±	-	±	-	-	-	-	-	-	-	+
31	+	+	+	+	+	+	+	-	-	-	-	-	-	-	+
32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
33	-	-	-	-	±	±	-	-	-	-	-	-	-	-	+
34	-	-	-	-	-	-	-	-	-	±	-	-	-	-	+
½	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 5. Hybridization of serotypes 1 and 9 *cps* genes with chromosomal DNA of other *S. suis* serotypes

Serotype	DNA probes *cps1E*	*cps1F*	*cps1G*	*cps1H*	*cps1I*	*cps9E*	*cps9F*	*cps9G*	*cps9H*	16rRNA
1	+	+	+	+	+	-	-	-	-	+
2	+	-	-	-	-	-	-	-	-	+
3	-	-	-	+	-	+	-	-	-	+
4	-	-	-	+	-	+	-	-	-	+
5	-	-	-	+	-	+	-	-	-	+
6	-	-	-	-	-	-	-	-	-	+
7	-	-	-	+	-	+	-	-	-	+
8	-	-	-	-	-	-	-	-	-	+
9	-	-	-	+	-	+	+	+	+	+
10	-	-	-	+	-	+	+	-	-	+
11	-	-	-	+	-	+	±	-	-	+
12	-	-	-	±	-	+	±	-	-	+
13	-	-	-	+	-	+	-	-	-	+
14	+	+	+	+	+	-	-	-	-	+
15	-	-	-	-	-	-	-	-	-	+
16	-	-	-	-	-	-	-	-	-	+
17	-	-	-	+	-	+	-	-	-	+
18	-	-	-	+	-	+	-	-	-	+
19	-	-	-	+	-	+	-	-	-	+
20	-	-	-	-	-	-	-	-	-	+
21	-	-	-	+	-	+	±	-	-	+
22	-	-	-	-	-	-	-	-	-	+
23	-	-	-	+	-	+	-	-	-	+
24	-	-	-	+	-	+	+	-	-	+
25	-	-	-	-	-	-	-	-	-	+
26	-	-	-	-	-	-	±	-	-	+
27	+	-	-	-	-	-	-	-	-	+
28	-	-	-	+	-	+	±	-	-	+
29	-	-	-	+	-	+	-	-	-	+
30	-	-	-	+	-	+	±	-	-	+
31	-	-	-	+	-	+	-	-	-	+
32	-	-	-	-	-	-	-	-	-	+
33	-	-	-	-	-	-	±	-	-	+
34	-	-	-	-	-	-	-	-	-	+
1/2	+	-	-	-	-	-	-	-	-	+

LEGENDS TO FIGURES

Fig.1.

Genetic organization of the *cps2* gene cluster.

(A) The arrows represent potential Orfs. Gene designations are indicated below the arrows.

(B) Physical map and genetic organization of the *cps2* locus on the chromosome of *S. suis* serotype 2.

Restriction sites are as follows: C: *Cla*I; E, *EcoR*I; H, *Hind*III; K, *Kpn*I; M, *Mlu*I; P, *Pst*I; S, *Sna*BI; Sa: *Sac*I; X, *Xba*I.

(C) The DNA fragments cloned in the various plasmids are indicated.

Fig.2.

Ethidium bromide stained agarose gel showing PCR products obtained with chromosomal DNA of *S.suis* strains belonging to the serotypes 1,2, ½, 9 and 14 and *cps2J, cps1I* and *cps9H* primer sets as described in Materials and Methods. (A) *cps1I* primers.

(B) *cps2J* primers and (C) *cps9H* primers. Lanes 1-3: serotype 1 strains; lanes 4-6: serotype 2 strains; lanes 7-9: serotype ½ strains; lanes 10-12: serotype 9 strains and lanes 13-15: serotype 14 strains.

(B) Ethidium bromide stained agarose gel showing PCR products obtained with tonsillar swabs collected from pigs carrying *S.suis* type 2, type 1 or type 9 strains and *cps2j, cps1I* and *cpsH* primer sets as described in Materials and Methods. Bacterial DNA suitable for PCR was prepared by using the multiscreen methods as described previously (20). (A) *cps1I* primers. (B) *cps2J* primers and (C) *cps9H* primers. Lanes 1-3: PCR products obtained with tonsillar swabs collected from pigs carrying *S.suis* type 1 strains; lanes 4-6: PCR products obtained with tonsillar swabs collected from pigs carrying *S.suis* type 2 strains; lanes 7-9: PCR procucts obtained with tonsillar swabs collected from pigs carrying *S.suis* type 9

strains; lanes 10-12: PCR products obtained with chromosomal DNA from serotype 9, 2 and 1 strains respectively; lane 13: negative control, no DNA present.

Figure 3
CPS2 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 4
CPS1 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

Figure 5
CPS9 nucleotide sequences and corresponding amino acid sequences from the open reading frames.

REFERENCES

1. **Arends, J. P., and H. C. Zanen.** 1988. Meningitis caused by *Streptococcus suis* in humans. Rev. Infect. Dis. **10**:131-137.

2. **Arrecubieta, C., E. Garcia, and R. Lopez.** 1995. Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. Gene **167**: 1-7

3. **Arrecubieta, C., R. Lopez, and E. Garcia.** 1994. Molecular characterization of *cap3A*, a gene from the operon required for the synthesis of the capsule of *Streptococcus pneumoniae* type 3: sequencing of mutations responsible for the unencapsulated phenotype and localization of the capsular cluster on the pneumococcal chromosome. J. Bacteriol. **176**: 6375-6383.

4. **Clifton-Hadley, F.A.** 1983. *Streptococcus suis* type 2 infections. Br. Vet. J. **139**:1-5.

5. **Charland, N., J. Harel, M. Kobisch, S. Lacasse, and M. Gottschalk.** 1998. Streptococcus suis serotype 2 mutants deficient in capsular expression. Microbiol. **144**:325-332.

6. **Cross, A. S.** 1990. The biological significance of bacterial encapsulation. Curr. Top. Microbiol. Immunol. **150**: 87-95.

7. **Elliott, S. D. and J. Y. Tai** . 1978. The type specific polysaccharide of *Streptococcus suis*. J. Exp. Med. **148**: 1699-1704.

8. **Feder, I., M. M. Chengappa, B. Fenwick, M. Rider and J. Staats.** 1994. Partial characterization of *Streptococcus suis* type 2 hemolysin. J. Clin. Microbiol. **32**:1256-1260.

9. **Gottschalk, M., R. Higgins, M. Jacques, M. Beaudoin, and J. Henrichsen.** 1991. Characterization of six new capsular types (23 through 28) of *Streptococcus suis*. J. Clin. Microbiol. **29**:2590-2594.

10. **Gottschalk, M., S. Lacouture, and J. D. Dubreuil.** 1995. Characterization of *Streptococcus suis* type 2 haemolysin.

Microbiology **141**:189-195.

11. **Gottschalk, M., A. Lebrun, M. Jacques, and R. Higgins.** 1990. Haemagglutination properties of *Streptococcus suis*. J. Clin. Microbiol. **28**: 2156-2158.

12. **Guidolin, A., J. M. Morona, R. Morona, D. Hansman, and J. C. Paton.** 1994. Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. 1994. Infect. Immun. **62**: 5384-5396.

13. **Guitierrez, C., and J. C. Devedjian.** 1989. Plasmid facilitating *in vitro* construction of PhoA fusions in *Escherichia coli.* Nucl. Acid. Res. **17**: 3999.

14. **Higgins, R., M. Gottschalk, M. Boudreau, A. Lebrun, and J. Henrichsen.** 1995. Description of six new capsular types (28 through 34) of *Streptococcus suis.* J. Vet. Diagn. Invest. **7**:405-406

15. **Jacobs, A. A., P. L. W. Loeffen, A. J. G. van den Berg, and P. K. Storm.** 1994. Identification, purification and characterization of a thiol-activated hemolysin (suilysin) of *Streptococcus suis*. Infect. Immun. **62**: 1742-1748.

16. **Jacques, M., M. Gottschalk, B. Foiry E.and R. Higgins.** 1990. Ultrastructural study of surface components of *Streptococcus suis*. J. Bacteriol. **172**:2833-2838.

17. **Klein P., M. Kanehisa and C. DeLisi.** 1985. The detection and classification of membrane spanning proteins. Biochim. Biophys. Acta. **851**: 468-476.

18. **Kolkman, M. A. B., D. A. Morrison, B. A. M. van der Zeijst, and P. J. M. Nuijten.** 1996. The capsule polysaccharide synthesis locus of *Streptococcus pneumoniae* serotype 14: identification of the glycosyl transferase gene *cps14E.* J. Bacteriol. 178: 3736-3541.

19. **Kolkman, M. A. B., W. Wakarchuk, P. J. M. Nuijten, and B. A. M. van der Zeijst.** 1997. Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding

glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. Mol. Microbiol. **26**: 197-208.

20. **Kolkman, M. A. B., B. A. M. van der Zeijst and P. J. M. Nuijten.** 1997. Functional analysis of glycosyltransferases encoded by the capsular polysaccharide biosynthesis locus of *Streptococcus pneumoniae* serotype 14. J. Biol. Chem. **272**: 1950219508.

21. **Konings, R. N. H., E. J. M. Verhoeven, and B. P H. Peeters.** 1987. pKUN vectors for the separate production of both DNA strands of recombinant plasmids. Methods Enzymol. **153**: 12-34.

22. **Korolik, V., B. N. Fry, M. R. Alderton, B. A. M. van der Zeijst, and P. J. Coloe.** 1997. Expression of *Campylobacter hyoilei* lipo-oligosaccharide (LOS) antigens in *Escherichia coli*. Microbiol. 143: 3481-3489.

23. **Leij, P. C. J., R. van Furth, and T. L. van Zwet.** 1986. In vitro determination of phagocytosis and intracellular killing of polymorphonuclear and mononuclear phagocytes. In Handbook of Experimental Immunology, vol. 2. Cellular Immunology, pp. 46.1-46.21. Edited by D. M. Weir, L. A. Herzenberg, C. Blackwell and L. A. Herzenberg. Blackwell Scientific Publications, Oxford.

24. **Lin, W. S., T. Cunneen, and C. Y. Lee.** 1994. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. J. Bacteriol. **176**: 7005-7016.

25. **Liu, D., A. M. Haase, L. Lindqvist, A.A. Lindberg, and P. R. Reeves.** 1993. Glycosyl transferases of O-antigen biosynthesis in *Salmonella enteritica*: Identification and characterization of transferase genes of group B, C2, and E1. J. Bacteriol. **175**: 3408-3413.

26. **Manoil, C., and J. Beckwith.** 1985. A transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82**: 8129-8133.

27. **Marsh, J. L., M. Erfle, and E. J. Wykes.** 1984. The pIC

plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene **32**:481-485.

28. **Miller, J.** 1972. Experiments in Molecular Genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

29. **Morona, J. K., R. Morona, and J. C. Paton.** 1997. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthesis pathway. Mol. Microbiol. **23**: 761-763.

30. **Muñoz, R., M. Mollerach, R. López and E. Garcia.** 1997. Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae;* formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. Mol. Microbiol. **25**: 79-92.

31. **Pearce B. J., Y. B. Yin, and H. R. Masure**. 1993. Genetic identification of exported proteins in *Streptococcus pneumoniae*. Mol. Microbiol. **9**: 1037-1050.

32. **Roberts, I. S.** 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Ann. Rev. Microbiol. **50**: 285-315.

33. **Rossbach, S., D. A. Kulpa, U. Rossbach, and F. J. de Bruin**. 1994. Molecular and genetic characterization of the rhizopine catabolism (mocABRC) genes of *Rhizobium meliloti* L5-30. Mol. Gen. Genet. **24**5: 11-24.

34. **Rubens, C. E., L. M. Heggen, R. F. Haft, and R. M. Wessels**. 1993. Identification of *cpsD,* a gene essential for type III capsule expression in group B streptococci. Mol. Microbiol. **8**: 843-855.

35. **Rubens, C. E., L. M. R. Wessels, L. M. Heggen, and D. L. Kasper**. 1987. Transposon mutagenesis of type III group B Streptococcus: correlation of capsule expression with virulence. Proc. Natl. Acad. Sci. USA **84**:7208-7212.

36. **Sambrook, J., E. F. Fritsch, and T. Maniatis**. 1989. Molecular cloning. A laboratory manual. Second edition. Cold

Spring Harbor Laboratory Press. Cold Spring Harbor. New York.

37. **Smith, H. E., U. Vecht, H. J. Wisselink, N. Stockhofe-Zurwieden, Y. Biermann, and M. A. Smits.** 1996. Mutants of *Streptococcus suis* types 1 and 2 impaired in expression of muramidase-released protein and extracellular protein induce disease in newborn germfree pigs. Infect Immun. **64**: 4409-4412.

38. **Smith, H. E., H. J. Wisselink, U. Vecht, A. L. J. Gielkens and M. A. Smits.** 1995. High-efficiency transformation and gene inactivation in *Streptococcus suis* type 2. Microbiol. **141**: 181-188.

39. **Sreenivasan, P. K., D. L. LeBlanc, L. N. Lee, and P. Fives-Taylor.** 1991. Transformation of *Actinobacillus actinomycetemcomitans* by electroporation, utilizing constructed shuttle plasmids. Infect. Immun. **59**: 4621-4627.

40. **Stringele F., J.-R. Neeser, and B. Mollet.** 1996. Identification and characterization of the *eps* (exopolysaccharide) gene cluster from *Streptococcus thermophilus* Sfi6. J. Bacteriol. **178**: 1680-1690.

41. **Stockhofe-Zurwieden, N., U. Vecht, H. J. Wisselink, H.van Lieshout, and H. E. Smith.** 1996. Comparative studies on the pathogenicity of different *Streptococcus suis* serotype 1 strains. In Proceedings of the 14th IPVS Congress. pp. 299.

42. **van Kranenburg, R., J. D. Marugg, I. I. van Swam, N. J. Willem and W. M. de Vos.** 1997. Molecular characterization of the plasmid-encoded *eps* gene cluster essential for exopolysaccharide biosynthesis in *Lactococcus lactis* Mol. Microbiol. **24**: 387-397.

43. **van Leengoed, L. A., E. M. Kamp, and J. M. A. Pol.** 1989. Toxicity of *Haemophilus pleuropneumoniae* to porcine lung macrophages. Vet. Microbiol. **19**: 337-349.

44. **van Leengoed, L. A. M. G., U. Vecht, and E. R. M. Verheyen.** 1987. *Streptococcus suis* type 2 infections in pigs in The Netherlands (part two). Vet Quart. 9, 111-117.

45. **Vecht, U., J. P. Arends, E. J. van der Molen, and L. A. M. G. van Leengoed.** 1989. Differences in virulence between two

strains of *Streptococcus suis* type 2 after experimentally induced infection of newborn germfree pigs. Am. J. Vet. Res. **50**:1037-1043.

46. **Vecht, U., L. A. M. G. van Leengoed, and E. R. M. Verheyen**. 1985. *Streptococcus suis* infections in pigs in The Netherlands (part one). Vet. Quart. **7**:315-321

47. **Vecht, U., H. J. Wisselink, M. L. Jellema, and H. E. Smith**. 1991. Identification of two proteins associated with virulence of *Streptococcus suis* type 2. Infect. Immun. **59**:3156-3162.

48. **Vecht, U., H. J. Wisselink, N. Stockhofe-Zurwieden, and H. E. Smith**. 1996. Characterization of virulence of the Streptococcus suis serotype 2 reference strain Henrichsen S 735 in newborn gnotobiotic pigs. Vet. Microbiol. **51**:125-136.

49. **Vecht, U., H. J. Wisselink, J. E. van Dijk, and H. E. Smith**. 1992. Virulence of *Streptococcus suis* type 2 strains in newborn germfree pigs depends on phenotype. Infect. Immun. **60**:550-556.

50. **Wagenaar, F., G. L. Kok, J. M. Broekhuijsen-Davies, and J. M. A. Pol**. 1993. Rapid cold fixation of tissue samples by microwave irradiation for use in electron microscopy. Histochemical J. **25**: 719-725.

51. **Wessels, M. R.and M. S. Bronze**. 1994. Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. Proc. Natl. Acad. Sci. USA **91**: 12238-12242.

52. **Wessels, M. R., A. E. Moses, J. B. Goldberg, and T. J. DiCesare**. 1991. Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. Proc. Natl. Acad. Sci. USA. **88**: 8317-8321.

53. **Yamane, K., M. Kumamano, and K.Kurita**. 1996. The 25^{o}-36^{o} region of the *Bacillus subtilis* chromosome: determination of the sequence of a 146 kb segment and identification of 113 genes. Microbiol. **142**: 3047-3056.

54. **Butler, J. C., R. F. Breiman, H. B. Lipman, J. Hofmann,**

and R. R. Facklam. 1995. Serotype distribution of *Streptococcus pneumoniae* infections among preschool children in the United States, 1978-1994: implications fro development of a conjugate vaccine. J. Infect. Dis. **171:** 885-889.

55. **Charland, N., M. Jacques, S. Lacoutre and M. Gottschalk.** 1997. Characterization and protective activity of a monoclonal antibody against a capsular epitope shared by *Streptococcus suis* serotypes 1, 2 and 1/2. Microbiol. **143:**3607-3614.

56. **Gottschalk, M., R. Higgins, M. Jacques, K. R. Mittal, and J. Henrichsen.** Description of 14 new capsular types of *Streptococcus suis.* J. Clin. Microbiol. **27:**2633-2636.

57. **Heath, P. J., B. W. Hunt, and J. P. Duff.** 1996. *Streptococcus suis* serotype 14 as a cause of pig disease in the UK. Vet. Rec. **2:**450-451.

58. **Hommez, J., L. A. Devrieze, J. Henrichsen, and F. Castryck.** 1986. Identification and characterization of *Streptococcus suis.* Vet. Microbiol. **16:**349-355.

59. **Killper-Balz, R., and K. H. Schleifer. 1987.** *Streptococcus suis* sp. nov. nom. rev. Int. J. Syst. Bacteriol. **37:**160-162.

60. **Kolkman, M. A. B., B. A. M. van der Zeijst, and P. J. M. Nuijten.** 1998. Diversity of capsular polysaccharide synthesis gene clusters in Streptococcus pneumoniae. Submitted for publication.

61. **Lee, J. C., S. Xu, A. Albus, and P. J. Livolsi.** 1994. Genetic analysis of type 5 capsular polysaccharide expression by *Staphylococcus aureus.* J. Bacteriol. **176:**4883-4889.

62. **Reek, F. H., M. A. Smits, E. M. Kamp, and H. E. Smith.** 1995. Use of multiscreen plates for the preparation of bacterial DNA suitable for PCR. BioTechniques **19:** 282-285.

63. **Sau, S., N. Bhasin, E. R. Wann, J. C. Lee, T. J. Foster, and C. Y. Lee.** 1997. The *Staphylococcus aureus* allelic genetic loci for serotype 5 and 8 capsule expression contain the type-specific genes flanked by common genes. Microbiol. **143:** 2395-2405.

64. **Sau, S., and C. Y. Lee.** 1996. Cloning of type 8 capsule genes and analysis of gene clusters for the production of different capsular polysaccharides in *Staphylococcus aureus*. J. Bacteriol. **178**: 2118-2126.

65. **Sau, S., and C. Y. Lee.** 1997. Molecular characterization and transcriptional analysis of type 8 capsule genes in *Staphylococcus aureus*. J. Bacteriol. **179**:1614-1621.

66. **Smith, H. E., M. Rijnsburger, N. Stockhofe-Zurwieden, H. J. Wisselink, U. Vecht, and M. A. Smits.** 1997. Virulent strains of *Streptococcus suis* serotype 2 and highly virulent strains of *Streptococcus suis* serotype 1 can be recognized by a unique ribotype profile. J. Clin. Microbiol. **35**:1049-1053.

67. **Yamazaki, M., L. Thorne, M. Mikolajczak, R. W. Armentrout, and T. J. Pollock.** 1996. Linkage of genes essential for synthesis of a polysaccharide capsule in *Sphingomonas* strain S88. J. Bacteriol. **178**:2676-2687.

68. **Zhang, L., A.Al-Hendy, P. Toivanen. and M. Skurnik.** 1993. Genetic organization and sequence of the *rfb* gene cluster of *Yersinia enterolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. Mol. Microbiol. **9**:309-321.

EPO - DG 1

22. 07. 1998

CLAIMS

1. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derived thereof.

2. A nucleic acid according to claim 1 encoding a *Streptococcus suis* serotype-specific central region, preferably encoding at least one enzyme or fragment thereof involved in polysaccharide biosynthesis.

3. A nucleic acid according to claim 1 or 2 hybridising to a nucleic acid encoding a gene derived from a *Streptococcus suis* serotype 1, 2 or 9 capsular gene cluster.

4. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 2 or a gene or gene fragment derived thereof, preferably as identified in Figure 3.

5. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 1 or a gene or gene fragment derived thereof, preferably as identified in Figure 4.

6. An isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* serotype 9 or a gene or gene fragment derived thereof, preferably as identified in Figure 5.

7. A nucleic acid probe or primer derived from a nucleic acid according to anyone of claims 1 to 6 allowing species or serotype specific detection of *Streptococcus suis*.

8. A probe or primer according to claim 7 provided with at least one reporter molecule.

9. A diagnostic test comprising a probe or primer according to claim 7 or 8.

10. A protein or fragment thereof encoded by a nucleic acid according to anyone of claims 1 to 6.

11. A protein or fragment according to claim 10 capable of polysaccharide biosynthesis.

12. A method to produce a *Streptococcus suis* capsular antigen comprising using a protein or fragment according to claim 11.

13. A *Streptococcus suis* capsular antigen obtainable by a method according to claim 12.

14. A vaccine comprising an antigen according to claim 13 and further comprising a suitable carrier or adjuvant.

15. A recombinant *Streptococcus suis* mutant provided with a modified capsular gene cluster.

16. A recombinant micro-organism comprising at least a part of a capsular gene cluster of *Streptococcus suis*.

17. A recombinant micro-organism according to claim 16 comprising a lactic acid bacterium.

18. A vaccine comprising a mutant according to claim 15 or a micro-organism according to claim 16 or 17.

EPO - DG 1

22. 07. 1998

ABSTRACT

The invention relates to *Streptococcus suis* infections of pigs, to vaccines directed against those infections and to tests for diagnosing *Streptococcus suis* infections.
The invention provides an isolated or recombinant nucleic acid encoding a capsular gene cluster of *Streptococcus suis* or a gene or gene fragment derived thereof. The invention furthermore provides a nucleic acid probe or primer allowing species or serotype specific detection of *Streptococcus suis*. The invention also provides a *Streptococcus suis* antigen and vaccine derived thereof.

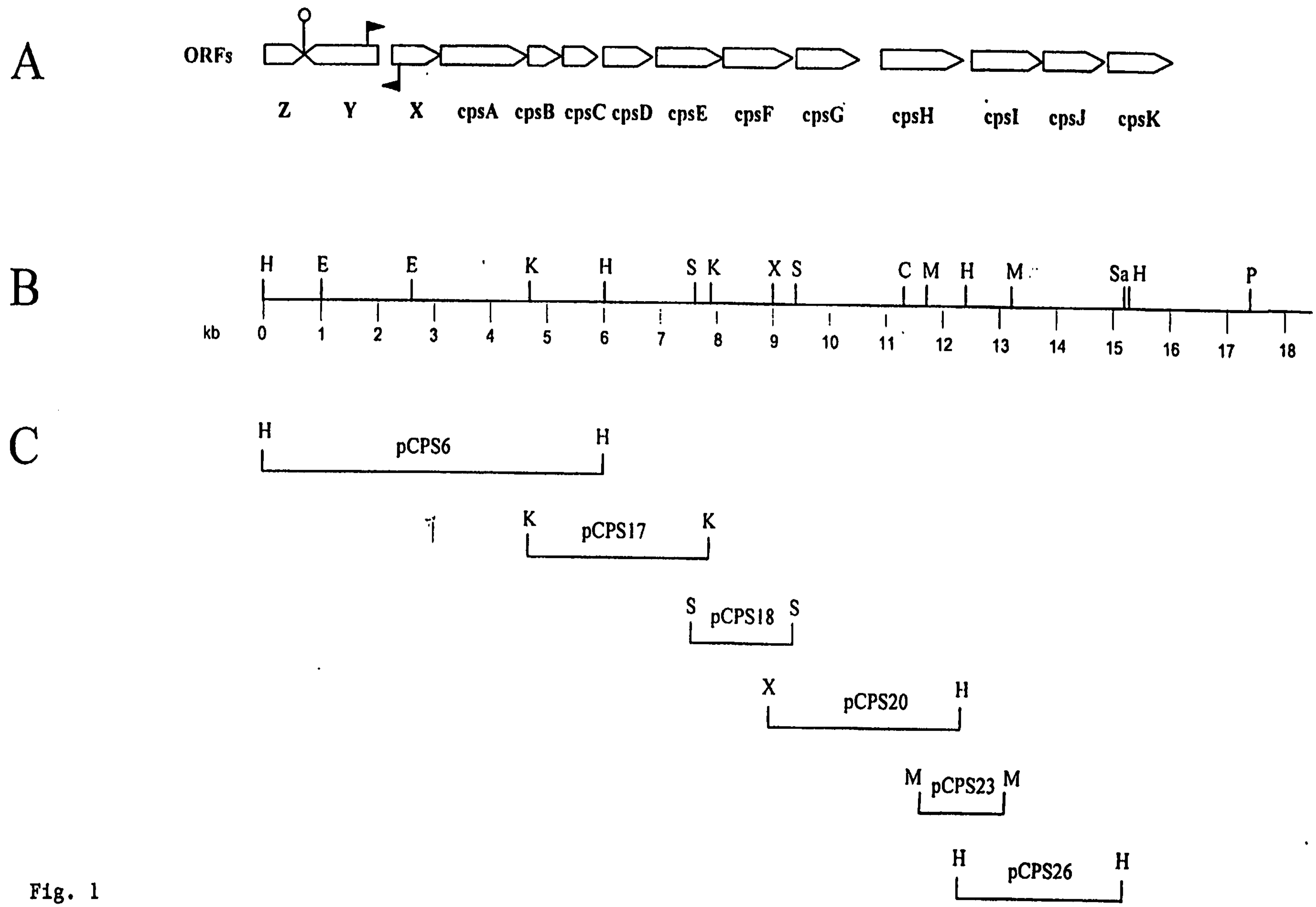
A
ORFs
Z
Y
X
cpsA
cpsB
cpsC
cpsD
cpsE
cpsF
cpsG
cpsH
cpsI
cpsJ
cpsK
B
H
E
E
K
H
S
K
X
S
C
M
H
M
Sa
H
P
kb
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
C
H pCPS6 H
K pCPS17 K
S pCPS18 S
X pCPS20 H
M pCPS23 M
H pCPS26 H

Fig. 1

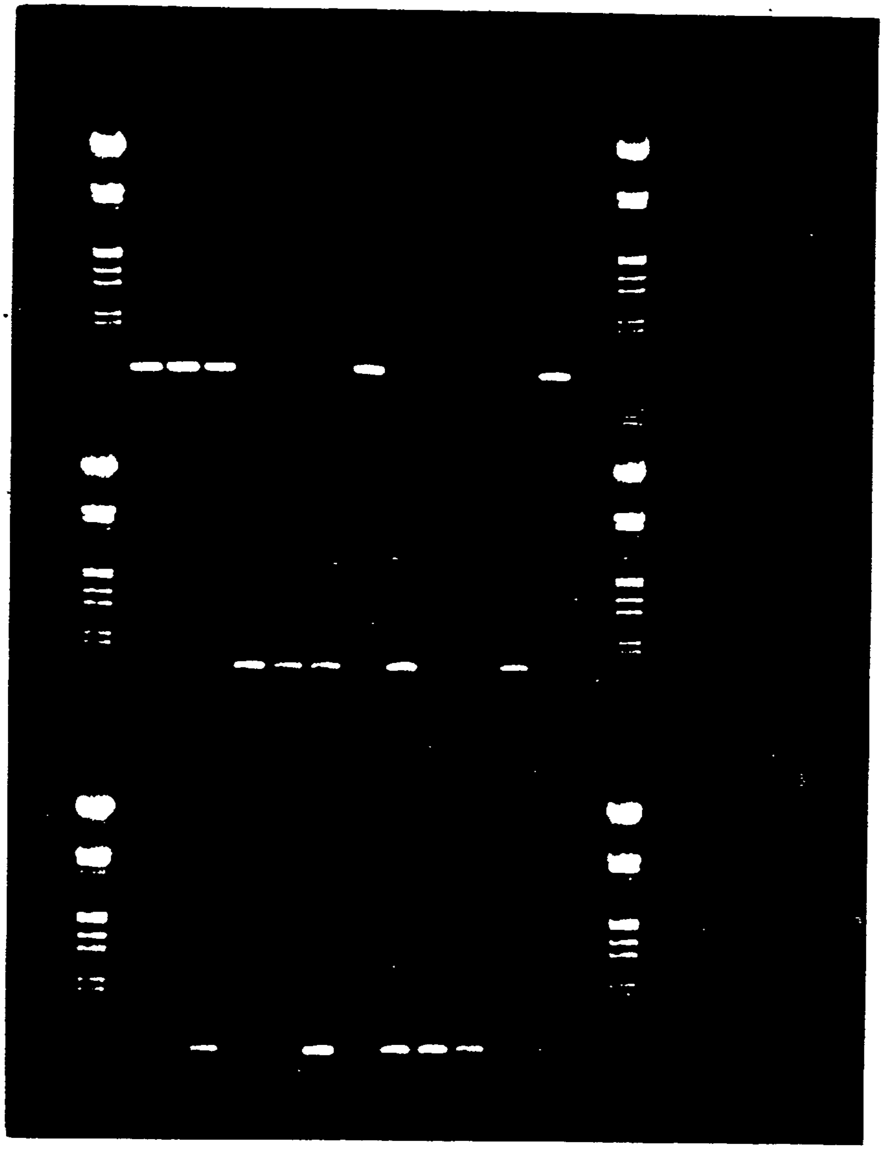

Figure 2

CPS2

AAGCTTGGAT ATTGATCACA TGATGGAGGT GATGGAAGCA TCTAAGTCTG CAGCGGGGTC GGCGTGCCCA
AGTCCGCAGG CTTATCAGGC AGCTTTTGAG GGAGCTGAGA
ACATTATCGT TGTGACGATT ACAGGTGGGC TATCGGGTAG TTTTAATGCG GCACGTGTAG CTAGGGATAT
GTATATCGAA GAGCATCCGA ATGTCAATAT CCATTTGATA
GATAGTTTGT CAGCCAGTGG GGAAATGGAT TTACTTGTAC ACCAAATCAA TCGCTTAATT AGTGCAGGAT
TAGATTTTCC ACAAGTAGTA GAAGCGATAA CTCACTATCG
GGAACACAGT AAGCTCCTCT TTGTTTTAGC GAAAGTTGAT AATCTTGTTA AGAATGGAAG ACTGAGCAAA
TTGGTAGGCA CTGTCGTTGG TCTTCTCAAT ATCCGTATGG
TTGGTGAGGC AAGTGCTGAA GGAAAATTAG AGTTGCTTCA AAAGGCGCGT GGTCATAAGA AATCTGTGAC
AGCAGCCTTT GAAGAAATGA AAAAAGCAGG CTATGATGGT
GGTCGAATTG TTATGGCCCA CCGCAACAAT GCTAAGTTCT TCCAACAATT CTCAGAGTTG GTAAAAGCAA
GTTTTCCAAC GGCTGTTATT GACGAAGTTG CAACATCAGG
TCTATGCAGT TTTTATGCTG AAGAAGGTGG ACTTTTGATG GGCTACGAAG TGAAAGCGTG ATTCACAGAG
TAATAATTTT GGGCTGTAAT TTCCGCTATA GAATAATCCC
CCTCTTCTTC TAAGTTCGAG GGGGATTGTT TGTATGAGAC TATTGGATTT CATTCATTCA AATATCTTAC
GAATTGCTCC AGTTTATCTG CAAAATCTTG TTCAAAGAAG
ATCTGTAAGA AATCAGCTTT CTGTCCGCTG AAATAATAAC ATTTTCCAAA CATGTGTTGG ATGCTAGGAG
AAAGAATCCC CTTGCTTAGC TGAAAGGTCA CGCTCCCCTT
TGGAATTCGA TACGGGATGT TTAAAGCGTA TTTCTCTAGA CAGTCTTTTA TTTTATTCCA TTGAGCGTGA
TAAATGTGAT GAAGATGCTG TGTGTTCCGC GCAAACATAC
CGTTATCAAT GTAGAGCGAG AGAGCTTTTT GCATGATAAG ATTGGTATCG TAGTCGATTA GACTCTTATG
TTTGATGAAG ATATCACGTA GCTGATTAGG AAGGCTGATT
GCACCGATTC GGAGGGCAGG AAAGAGTGTC GGTGTAAAAG ATTTTATATA GATGACGCGA TTATCTGTAT
CAAGATAGTG TAAAGGTAGG CTATGACTAG AGTCGAAATC
TGCTAAATAG TCATCCTCAA TGATGTAGAC ATCGTATTGC TTTGCTAATT TTACGATGGC TGTTTTTGTT
GCTATATCAT AGGTTGAACC GAGAGGGTTG TGCAAGCGAG
GAATTGTGTA GAAAAACTTA ATTTTTCCAG TTTGGAAGAT ACTTTCCAAT TCTTCTAGGT CAATTCCATC
TAAATTCCGT TCAATTGTTT GATAGGGGAT TCCTTGATGT
CGAATGAGCT CTATCATTCG TGAATAGGTA GGGTTCTCTA TCAAGATTTC CGTTTTTCCA GCCAAGGTTT
CCATTTGTGT GAGAATATAT AGAGCTTGTT GACTACCAGC
TGTGATAACC AGCTGGTCTT TTTTTGTATA GACATGATAG TCCATTAACA GACTTTGAAC GGAGGAAATC
AATTCTGCCA ATCCCTCTTG CTGGTGATAG TAGTTGAATA
GGTAATTTTC CCGCCCAATA AGACTTTCTT TTAGACAAAT CCGAAAATCT TCATAGGTAA TTCTTGAAAG
TCTGTAGGAT TGAGCTCTAC AGGTATGGTC TTGGAAATCT
CTATCCTCTA AGATATAATA ACCGCTTTTT TCGACAGCGT AGATCTTATT TTGGTATTTT AATTCCAACA
TAGCCTTTTG GACAGTGTCT TTGCTACAAT GATATTGCTC
GCGGAGTTGA CGGATAGAAG GTAATTTCTC TCCACGTTTG AATCGATGTT CCTCTATTCC AGTCAAAATA
TCTTGGATGA TAACTTGATA TTTTTTCATC TAGGTCCCCT
TTTTTATAGA CTATGTTACT AGCTAGTATA TAGAAAAAAT TGAAGAAAGA CAATATATGA ATAATGGGGT
TGAGGTTCAG GAATTAAGCT ACTCTATGGT ATAATTAAGT
GATGAAAATA ATTATACCTA ATGCAAAAGA AGTAAATACA AATCTAGAGA ATGCCTCGTT TTATCTCCTG
TCTGATCGAA GCAAGCCGGT GCTGGATGCC ATAAGTCAAT
TTGATGTAAA AAAGATGGCT GCCTTTTATA AATTGAATGA AGCAAAGGCT GAGTTAGAAG CTGACCGTTG
GTATCGAATC AGGACAGGTC AAGCAAAAAC CTATCCAGCC
TGGCAGTTAT ATGATGGTCT CATGTATCGT TATATGGATA GGCGAGGTAT AGATTCGAAA GAAGAAAATT
ATTTACGTGA CCACGTTCGT GTAGCGACAG CCTTATACGG
ATTGATTCAT CCTTTTGAAT TCATTTCACC TCACCGCTTA GATTTTCAAG GGAGCTTAAA GATAGGCAAT
CAGTCTTTGA AACAGTACTG GCGACCGTAT TATGACCAAG
AAGTTGGTGA TGATGAACTG ATTCTCTCAC TGGCTTCGTC AGAATTTGAG CAGGTGTTTT CTCCNCAGAT
TCAGAAAAGA TTAGTTAAAA TTCTTTTCAT GGAAGAAAAA
GCAGGTCAGC TAAAAGTTCA CTCGACTATA TCAAAAAAAG GCAGAGGAAG ATTGCTGTCC TGGTTGGCTA
AGAACAATAT TCAGGAATTA TCGGACATTC AAGATTTTAA
GGTGGATGGC TTTGAATATT GTACTTCCGA ATCAACGGCA AACCAACTTA CCTTCNTACG ATCAATAAAA
ATGTGAAATT ATGAAAAAGA TAACGTTTTC CAGCGCTAAA
AAGGGTAGAA AATATTAAT TTCTATGATA TAATGGATGC GTTATAGGTA AAAGTCTAGG AAGGTTGTTT
ATGAAAAAGA GAAGCGGACG AAGTAAGTCG TCCAAGTTCA
AATTGGTAAA TTTTGCGCTT TTGGGACTTT ATTCCATTAC TCTATGTTTG TTCTTAGTGA CCATGTATCG
CTATAACATC CTAGATTTCC GGTATTTAAA CTATATTGTG
ACGCTTTTGC TAGTAGGAGT GGCAGTATTG GCTGGATTAT TGATGTGGCG TAAGAAAGCG CGCATATTTA
CAGCGCTCTT ACTTGTTTTT TCACTGGTCA TCACGTCTGT
TGGGATCTAT GGAATGCAAG AAGTTGTAAA ATTTTCAACA CGACTAAATT CAAATTCGAC ATTTTCAGAA

Fig. 3

TATGAAATGA GTATCCTTGT CCCAGCAAAT AGTGATATTA
CGGACGTTCG TCAGCTTACT AGTATCCTTG CTCCAGCCGA ATACGACCAA GATAACATCA CCGCTTTATT
GGATGACATA TCCAAAATGG AATCTACTCA ACTAGCAACT
AGCCCCGGGA CTTCTTACCT GACAGCATAT CAATCTATGT TGAATGGCGA GAGTCAAGCG ATGGTGTTCA
ACGGAGTTTT TACCAATATT TTAGAAAATG AAGATCCAGG
CTTTTCTTCA AAAGTGAAAA AAATATATAG TTTCAAAGTG ACTCAGACTG TTGAAACAGC TACTAAGCAG
GTGAGTGGAG ATAGCTTTAA TATCTATATT AGTGGTATTG
ATGCTTATGG ACCGATTTCT ACGGTCTCTC GTTCAGATGT CAATATCATT ATGACTGTCA ATCGTGCGAC
ACATAAGATT TTATTGACAA CTACTCCACG AGATTCATAC
GTTGCTTTCG CAGATGGCGG GCAAAATCAA TACGATAAAC TAACACATGC TGGTATTTAC GGTGTCAATG
CTTCTGTGCA CACCTTAGAA AATTTTTATG GGATTGACAT
TAGCAATTAT GTGCGGTTGA ACTTCATTTC CTTCCTTCAA TTAATCGACT TGGTGGGTGG AATTGATGTA
TATAACGATC AAGAATTTAC AAGTTTACAT GGGAATTATC
ATTTCCCTGT TGGACAAGTT CATTTAAACT CAGACCAAGC ATTAGGCTTC GTTCGAGAGC GCTACTCTTT
AACAGGGGGT GACAATGACC GTGGTAAAAA CCAGGAAAAA
GTGATTGCTG CCTTGATTAA AAAGATGAGT ACGCCAGAGA ATCTAAAAAA TTACCAGGCA ATCCTATCTG
GATTGGAAGG CTCAATTCAA ACGGATTTGA GCTTAGAAAC
GATTATGAGT TTAGTGAATA CCCAACTAGA ATCAGGAACA CAATTTACAG TAGAGTCACA AGCATTGACA
GGAACAGGAC GCTCAGACTT ATCTTCTTAT GCGATGCCTG
GATCACAACT TTATATGATG GAAATTAACC AAGATAGTCT GGAGCAATCA AAGGCAGCGA TTCAGTCCGT
ACTTGTTGAA AAATAAAGAT TTTAGGAGAA AATATGAACA
ATCAAGAAGT AAATGCAATC GAAATCGATG TTTTATTCTT ACTAAAAACA ATTTGGAGAA AGAAATTTTT
AATTCTCTTA ACTGCAGTGT TGACTGCGGG GTTGGCATTT
GTCTACAGTA GTTTTTTAGT GACACCTCAA TATGACTCCA CTACCCGTAT CTATGTAGTG AGTCAAAATG
TTGAAGCCGG TGCGGGCTTG ACTAACCAAG AGTTACAAGC
GGGTACCTAT TTGGCAAAAG ACTATCGGGA AATTATCCTA TCACAAGATG TNTTGACACA AGTAGCAACG
GAATTGAATC TGAAAGAGAG TTTGAAAGAA AAAATATCAG
TTTCTATTCC TGTTGATACT CGTATCGTTT CTATTTCTGT GCGTGATGCG GATCCAAATG AAGCGGCACG
TATTGCAAAT AGCCTTCGCA CCTTTGCAGT GCAAAAGGTT
GTTGAGGTCA CCAAGGTAAG CGATGTGACG ACACTTGAAG AAGCAGTCCC AGCGGAAGAA CCAACCACTC
CAAATACAAA ACGAAATATC TTGCTTGGTT TATTAGCTGG
AGGTATCTTG GCAACAGGTC TTGTACTGGT TATGGAGGTT TTGGATGACC GTGTAAAACG TCCTCAGGAC
ATCGAAGAGG TAATGGGATT GACATTGCTA GGTATAGTAC
CAGATTCGAA GAAATTAAAA TAGGAGAACA ATATGGCGAT GTTAGAAATT GCACGTACAA AAAGAGAGGG
AGTAAATAAA ACCGAGGAGT ATTTCAATGC TATCCGTACC
AATATTCAGC TTAGCGGAGC AGATATTAAG GTTGTTGGTA TTACCTCTGT TAAATCGAAT GAAGGTAAGA
GTACAACTGC GGCTAGTCTC GCTATTGCCT ATGCTCGTTC
AGGTTATAAG ACCGTCTTGG TGGATGCAGA TATCCGAAAT TCAGTCATGC CTGGTTTCTT CAAGCCAATT
ACAAAGATTA CAGGTTTGAC GGATTACCTA GCAGGGACAA
CAGACTTGTC TCAAGGATTA TGCGATACAG ATATTCCAAA CTTGACCGTA ATTGAGTCAG GAAAGGTTTC
TCCCAACCCT ACTGCCCTTT TACAAAGTAA GAATTTTGAA
AATCTACTTG CGACTCTTCG TCGCTATTAT GATTATGTTA TCGTTGACTG TCCACCATTA GGACTGGTAA
TTGATGCAGC TATCATTGCA CAAAAATGTG ATGCGATGGT
TGCAGTAGTA GAAGCAGGCA ATGTTAAGTG CTCATCTTTG AAAAAAGTAA AAGAGCAGTT GGAACAAACA
GGCACACCGT TCTTAGGCGT TATCTTGAAC AAATATGATA
TTGCCACTGA GAAGTATAGT GAATACGGAA ATTACGGCAA AAAAGCCTAA TTTCTCAGAT AACATAAGTT
TGATAAGTAG GTATTAATAT GATTGATATC CATTCGCATA
TCATATTTGG TGTGGATGAC GGTCCCAAAA CTATTGAAGA GAGCCTGAGT TTGATAAGCG AAGCTTATCG
TCAAGGTGTT CGCTATATCG TAGCGACATC TCATAGACGA
AAAGGGATGT TTGAAACACC AGAAAAAATC ATCATGATTA ACTTTCTTCA ACTTAAAGAG GCAGTAGCAG
AAGTTTATCC TGAAATACGA TTGTGCTATG GTGCTGAATT
GTATTATAGT AAAGATATCT TAAGCAAACT TGAAAAAAAG AAAGTACCAA CACTTAATGG CTCGTGCTAT
ATTCTCTTGG AGTTCAGTAC GGATACTCCT TGGAAAGAGA
TTCAAGAAGC AGTGAACGAA ATGACGCTAC TTGGGCTAAC TCCCGTACTT GCCCATATAG AGCGTTATGA
TGCTCTGGCA TTTCAGTCAG AGAGAGTAGA AAAGCTAATT
GACAAGGGAT GCTACACTCA GGTAAATAGT AACCATGTGT TGAAGCCTGC TTTAATTGGC GAACGAGCAA
AAGAATTTAA AAAACGTACT CGATATTTTT TAGAGCAGGA
TTTAGTACAT TGTGTTGCTA GCGATATGCA TAATTTATAT AGTAGACCTC CGTTTATGAG GGAGGCGTAT
CAGCTTGTAA AAAAAGAGTA TGGTGAGGAT AGAGCGAAGG
CTTTGTTCAA GAAAAATCCT TTGTTGATAT TGAAAAATCA AGTACAGTAA CCTCATAGAA ATAGTGGAGG
AGCTATGAAT ATTGAAATAG GATATCGCCA AACGAAATTG

Fig. 3 cont.

GCATTGTTTG ATATGATAGC AGTTACGATT TCTGCAATCT TAACAAGTCA TATACCAAAT GCTGATTTAA
ATCGTTCTGG AATTTTTATC ATAATGATGG TTCATTATTT
TGCATTTTTT ATATCTCGTA TGCCGGTTGA ATTTGAGTAT AGAGGTAATC TGATAGAGTT TGAAAAAACA
TTTAACTATA GTATAATATT TGTAATTTTT CTTATGGCAG
TTTCATTTAT GTTAGAGAAT AATTTCGCAC TTTCAAGACG TGGTGCCGTG TATTTCACAT TAATAAACTT
CGTTTTGGTA TACCTATTTA ACGTAATTAT TAAGCAGTTT
AAGGATAGCT TTCTATTTTC GACAACCTAT CAAAAAAAGA CGATTCTAAT TACAACGGCT GAACTATGGG
AAAATATGCA AGTTTTATTT GAATCAGATA TACTATTTCA
AAAAAATCTT GTTGCATTGG TAATTTTAGG TACAGAAATA GATAAAATTA ATTTACCATT ACCGCTCTAT
TATTCTGTTG AAGAAGCTAT AGGGTTTTCA ACAAGGGAAG
TGGTCGACTA CGTCTTTATA AATTTACCAA GTGAATATTT TGACTTAAAG CAATTAGTTT CAGACTTTGA
GTTGTTAGGT ATTGATGTAG GCGTTGATAT TAATTCATTC
GGTTTTACTG TGTTGAAGAA TAAAAAAATC CAAATGCTAG GTGACCATAG CATCGTCACT TTTTCCACAA
ATTTTTATAA GCCTAGTCAC ATCTGGATGA AACGACTTTT
AGATATACTT GGAGCAGTAG TCGGGTTAAT TATTAGTGGT ATAGTTTCTA TTTTGTTAAT TCCAATTATT
CGTAGAGATG GTGGGCCAGC CATTTTTGCT CAGAAACGAG
TTGGACAGAA TGGACGCATA TTTACATTCT ACAAGTTTCG TTCGATGTTT GTTGATGCCG AGGTACGTAA
GAAAGAATTA ATGGCTCAAA ACCAGATGCA AGGTGGGATG
TTCAAAATGG ACAACGATCC TAGAATTACT CCAATTGGAC ACTTCATACG AAAAACAAGT TTAGATGAGT
TACCACAATT TTATAATGTT CTAATTGGAG ATATGAGTCT
AGTCGGTACC CGTCCGCCTA CAGTTGATGA ATTTGAAAAA TATACTCCTA GTCAAAAGAG AAGATTGAGT
TTTAAACCAG GGATTACAGG TCTTTGGCAA GTGAGCGGAA
GAAGTGATAT CACAGATTTT AATGAAGTCG TTAGGCTGGA CCTAACATAC ATTGATAATT GGACCATCTG
GTCAGACATT AAGATTTTAT TGAAGACAGT GAAAGTTGTA
TTGTTGAGAG AGGGAGGTCA GTAAGACTCC TTTAAAACAA AGAATAGTAG TAGGGGATAT GAGAACAGTT
TATATTATTG GTTCAAAAGG AATACCAGCA AAGTATGGTG
GTTTCGAGAC TTTCGTAGAA AAATTAACTG AGTATCAGAA AGATAAATCA ATTAATTATT TTGTTGCATG
TACAAGAGAA AATTCAGCAA AATCAGATAT TACAGGAGAA
GTTTTTGAAC ATAATGGAGC AACATGTTTT AATATTGATG TGCCAAATAT TGGTTCAGCA AAAGCCATTC
TTTATGATAT TATGGCTCTC AAGAAATCTA TTGAAATTGC
CAAAGATAGA AATGATACCT CTCCAATTTT CTACATTCTT GCTTGTCGGA TTGGTCCTTT CATTTATCTT
TTTAAGAAGC AGATTGAATC AATTGGAGGT CAACTTTTCG
TAAACCCAGA CGGTCATGAA TGGCTACGTG AAAAGTGGAG TTATCCCGTC CGACAGTATT GGAAATTTTC
TGAGAGTTTG ATGTTAAAAT ACGCTGATTT ACTAATTTGT
GATAGCAAAA ATATTGAAAA ATATATTCAT GAAGATTATC GAAAATATGC TCCTGAAACA TCTTATATTG
CTTATGGAAC AGACTTAGAT AAATCACGCC TTTCTCCGAC
AGATAGTGTA GTACGTGAGT GGTATAAGGA GAAGGAAATT TCAGAAAATG ATTACTATTT GGTTGTTGGA
CGATTTGTGC CTGAAAATAA CTATGAAGTA ATGATTCGAG
AGTTTATGAA ATCATATTCA AGAAAAGATT TTGTTTTGAT AACGAATGTA GAGCATAATT CCTTTTATGA
GAAATTGAAA AAAGAAACAG GGTTCGATAA AGATAAGCGT
ATAAAGTTTG TTGGAACAGT CTATAATCAG GAGCTGTTAA AATATATTCG TGAAAATGCA TTTGCTTATT
TTCATGGTCA CGAGGTTGGA GGAACGAACC CATCTTTACT
TGAAGCACTT TCTTCTACTA AACTAAATCT TCTTCTAGAT GTGGGCTTTA ATAGAGAAGT AGGGGAAGAA
GGAGCGAAAT ACTGGAATAA AGATAATCTT CACAGAGTTA
TTGACAGTTG TGAGCAATTA TCACAAGAAC AAATTAATGA TATGGATAGT TTATCAACAA AACAAGTCAA
AGAAAGATTT TCTTGGGATT TTATTGTTGA TGAGTATGAG
AAGTTGTTTA AAGGATAAGT TATGAAAAAG ATTCTATATC TCCATGCTGG AGCAGAATTA TATGGGGCAG
ATAAGGTTCT CTTGGAACTT ATAAAAGGCT TAGATAAGAA
TGAATTTGAA GCGCATGTTA TCCTACCTAA TGATGGAGTC CTAGTGCCAG CATTAAGAGA AGTTGGTGCG
CAAGTTGAAG TTATTAACTA TCCAATTCTA CGTAGGAAAT
ATTTTAATCC AAAAGGGATT TTTGACTACT TCATATCATA TCATCACTAT TCTAAACAGA TTGCTCAATA
TGCCATAGAA AATAAGGTTG ACATAATTCA CAATAATACT
ACCGCTGTCT TAGAAGGCAT TTATCTGAAG CGAAAACTCA AATTACCTTT GTTGTGGCAT GTTCATGAGA
TTATTGTCAA ACCTAAATTC ATCTCTGATT CGATCAATTT
TTTAATGGGG CGTTTTGCTG ATAAGATTGT GACAGTTTCA CAGGCTGTGG CAAACCATAT AAAACAATCA
CCTCATATCA AAGATGACCA AATCAGTGTA ATCTACAATG
GGGTAGATAA TAAAGTGTTT TATCAGTCCG ATGCTCGGTC TGTTCGAGAA AGATTTGACA TTGACGAAGA
GGCTCTTGTC ATTGGTATGG TCGGTCGAGT CAATGCGTGG
AAAGGACAAG GAGATTTTTT AGAAGCAGTT GCTCCTATAC TCGAACAGAA TCCAAAAGCT ATCGCCTTTA
TAGCAGGAAG TGCTTTTGAA GGAGAAGAGT GGCGAGTAGT
AGAATTAGAA AAGAAGATTT CTCAATTAAA GGTCTCTTCT CAAGTCAGAC GAATGGATTA TTATGCAAAT

Fig. 3 cont.

ACCACTGAAT TATATAATAT GTTTGATATT TTTGTACTTC
CAAGTACTAA TCCAGACCCT CTACCAACGG TTGTACTAAA AGCAATGGCA TGCGGTAAAC CTGTTGTCGG
TTACCGACAT GGTGGTGTTT GTGAGATGGT GAAAGAAGGT
GTTAACGGTT TCTTAGTCAC TCCGAACTCA CCGTTAAATT TATCAAAAGT AATTCTTCAG TTATCGGAAA
ATATAAATCT CAGAAAAAAA ATTGGTAATA ATTCTATAGA
ACGTCAAAAA GAACATTTTT CGTTAAAAAG CTATGTAAAA AATTTTTCGA AAGTCTACAC CTCCCTCAAA
GTATACTGAT TGGCTGAAGT GAATGCTTTA GTATAGCGAT
TTATCGTATT CTCATTCGAT AAAACAAATG TTCAGAAACA GTTATAAGTT ATTTCTAAAG GGCACCTCTA
TAAACTCCCA AAATTGCGAA TTTGGAGTTA CGAAAGCCTT
GTTAAATCAA CATTTTAAAT TTTAGAAAAT TAGTTTTTAG AGCTCCCCTA AAATAGAAGA TAACAGAAGG
GAGCCTTCAA AAACTTCATT TTTAATTGGA TTGTAGAAAA
ACTGTTAAAT CAATATTTAG ATTTTTAGGA GTTCAGTTTT TGGGGGGAGA GCTTAATAAT CTATGCACTA
TATTTCGAAA AATATATGGT GTAAAATCAG AACTGATGGT
CGTGGCAAAA AAGAGAATGA GGAATTTATG AAAATTATTT CTTTTACAAT GGTTAATAAC GAAAGTGAGA
TAATAGAGTC ATTTATACGG TATAATTATA ACTTTATTGA
CGAGATGGTC ATTATTGATA ATGGTTGTAC AGATAACACG ATGCAAATTA TTTTTAATTT GATTAAAGAG
GGATATAAAA TATCCGTATA TGATGAGTCT TTAGAGGCAT
ATAATCAGTA TCGACTTGAT AATAAATATC TAACGAAAAT AATTGCTGAA AAAAATCCAG ATTTGATAAT
ACCTTTGGAT GCGGATGAAT TTTTAACAGC CGATTCAAAT
CCACGGAAAC TTTTGGAACA ACTGGACTTA GAAAAGATAC ATTATGTGAA TTGGCAATGG TTTGTTATGA
CTAAAAAAGA TGATATTAAT GATTCGTTTA TACCACGTAG
AATGCAATAT TGTTTTGAAA AACCTGTTTG GCATCATTCT GATGGTAAAC CAGTTACTAA ATGTATAATT
TCCGCTAAGT ATTACAAAAA AATGAATTTA AAGCTATCGA
TGGGACATCA CACTGTTTTT GGTAACCCAA ATGTAAGGAT AGAACATCAT AATGATTTGA AATTTGCACA
TTATCGAGCT ATTAGCCAAG AGCAATTAAT TTATAAAACA
ATTTGTTACA CTATTCGCGA TATTGCTACT ATGGAGAACA ATATCGAAAC AGCTCAAAGA ACAAATCAGA
TGGCGCTCAT TGAATCTGGC GTGGATATGT GGGAAACGGC
GAGAGAAGCC TCTTATTCAG GTTATGATTG TAATGTTATA CATGCACCAA TTGATTTAAG TTTTTGTAAA
GAAAATATTG TAATAAAATA TAACGAACTA TCCAGAGAAA
CAGTAGCAGA ACGCGTGATG AAAACGGGAA GAGAAATGGC TGTTCGTGCA TATAATGTGG AGCGAAAACA
AAAAGAAAAG AAATTTCTAA AACCTATTAT ATTTGTATTA
GATGGGTTAA AAGGAGATGA GTATATTCAT CCCAATCCAT CAAATCATTT GACGATCTTA ACTGAAATGT
ATAACGTCAG AGGCTTACTT ACCGATAATC ACCAAATTAA
ATTTCTCAAA GTTAATTATA GATTAATTAT AACTCCAGAT TTTGCTAAGT TTTTACCGCA TGAATTTATT
GTTGTACCAG ATACCTNGGA TATAGAGCAA GTTAAAAGCC
AGTATGTTGG TACAGGTGTA GACTTGTCAA AGATTATTTC TTTAAAAGAG TATCGAAAAG AGATAGGCTT
TATTGGTAAT TTGTATGCGC TTTTAGGATT TGTTCCGAAT
ATGCTCAATA GAATTTATCT ATATATTCAG AGAAACGGTA TTGCAAACAC TATTATAAAA ATCAAGTCGA
GATTGTGAGA GTTGTTTACT TTTATTTGTA ATTTTAAAAG
TAATGCAGGC AGATAGGAGA AAAACGTTTG GAAAAATGAG AATAAGAATT AATAATTTGT TTTTTGTTGC
CATAGCGTTT ATGGGCATAA TTATTAGTAA TTCGCAAGTT
GTTCTAGCGA TAGGCAAAGC TTCTGTGATT CAGTATCTAT CTTATTTAGT TTTGATTTTA TGTATAGTTA
ATGATTTATT AAAAAATAAC AAACATATTG TAGTTTATAA
ATTAGGGTAT TTGTTTCTTA TTATATTTTT ATTTACTATC GGAATATGTC AGCAAATTCT TCCTATAACA
ACTAAAATAT ATTTATCAAT TTCAATGATG ATTATTTCAG
TTTTAGCAAC GTTGCCAATA AGTTTGATAA AAGATATTGA TGATTTTAGA CGGATTTCAA ATCATTTGTT
ATTCGCTCTT TTTATAACTT CGATATTAGG AATAAAGATG
GGGGCAACGA TGTTCACGGG GGCAGTAGAA GGTATCGGTT TTAGTCAGGG TTTTAATGGA GGATTGACGC
ATAAGAACTT TTTTGGAATA ACTATTTTAA TGGGGTTCGT
ATTAACTTAC TTGGCGTATA AGTATGGTTC CTATAAAAGA ACGGATCGTT TTATTTTAGG ATTAGAATTG
TTTTTGATTC TTATTTCAAA CACACGCTCA GTTTATTTAA
TACTATTGCT TTTTCTATTT CTTGTTAATC TTGACAAAAT CAAAATAGAA CAAAGACAAT GGAGTACGCT
TAAATATATT TCCATGCTAT TTTGTGCTAT TTTTTTATAC
TATTTCTTTG GTTTTTTAAT AACACATAGT GATTCTTACG CTCATCGCGT TAATGGTCTT ATTAATTTTT
TTGAGTATTA TAGAAATGAT TGGTTCCATC TAATGTTTGG
TGCAGCGGAT TTGGCATATG GGGATTTAAC TTTAGACTAT GCTATAAGGG TTAGACGCGT TTTAGGTTGG
AATGGAACGC TTGAAATGCC CTTACTGAGT ATTATGTTAA
AAAATGGTTT TATCGGTCTG GTAGGGTATG GGATTGTTTT ATATAAACTT TATCGTAATG TAAGAATATT
AAAAACAGAT AATATAAAAA CAATAGGAAA GTCTGTATTT
ATCATTGTAG TCCTATCTGC AACAGTAGAA AATTATATTG TAAATTTAAG TTTTGTATTT ATGCCAATAT
GTTTTTGTTT ATTAAATTCT ATATCTACTA TGGAATCAAC

Fig. 3 cont.

TATTAACAAA CAACTGCAAA CATAAATTGG CAGGAATAGA GTTTTGAGTT GCTATTAATT TGGTAGAGCA
TATGTTCTAT AGGTGGCAAG ATAAAGATAG TATTTTTTAC
ATGATGATTT TTATGATAGC AAAGCAAGTT ACGGCATAAA AGGAATTAGA GGATGGAAAA AGTCAGCATT
ATTGTACCTA TTTTTAATAC GGAAAAGTAC TTAAGAGAGT
GTTTAGATAG CATTATTTCC CAATCGTATA CTAATCTAGA GATTCTTTTG ATAGATGACG GTTCTTCAGA
TTCATCAACG GATATATGTT TGGAATACGC AGAGCAAGAT
GGTAGAATAA AACTTTTCCG GTTACCAAAT GGTGGTGTTT CAAACGCAAG GAATTACGGT ATCAAAAATA
GCACAGCAAA TTATATTATG TTTGTAGATT CTGATGATAT
TGTTGACGGC AACATTGTTG AGTCCTTATA CACCTGTTTA AAAGAGAATG ATAGTGATTT GTCGGGAGGG
TTACTTGCTA CTTTTGATGG AAATTATCAA GAATCTGAGC
TGCAAAAGTG TCAAATTGAT TTGGAAGAGA TAAAAGAGGT GCGAGACTTA GGAAATGAAA ATTTTCCCAA
TCATTATATG AGCGGTATCT TTAATAGCCC TTGTTGCAAA
CTTTATAAGA ATATATATAT AAACCAAGGT TTTGACACTG AACAGTGGTT AGGAGAGGAC TTATTATTTA
ATCTAAATTA TTTAAAGAAT ATAAAAAAAG TCCGCTATGT
TAACAGAAAT CTTTATTTTG CCAGAAGAAG TTTACAAAGT ACTACAAATA CGTTTAAATA TGATGTTTTT
ATTCAATTAG AAAATTTAGA AGAAAAAACT TTTGATTTGT
TTGTTAAAAT ATTTGGTGGA CAATATGAAT TTTCTGTTTT TAAAGAGACG CTACAGTGGC ATATTATTTA
TTATAGCTTA TTAATGTTCA AAAATGGAGA TGAATCGCTT
CCAAAGAAAT TGCATATATT TAAGTATTTA TACAATAGGC ATTCTTTAGA TACTCTAAGT ATTAAACGAA
CGTCCTCTGT TTTTAAAAGA ATATGTAAAT TAATTGTTGC
TAATAATTTG TTTAAAATTT TTTTAAATAC TTTAATTAGG GAAGAAAAAA ATAATGATTA ACATTTCTAT
CATCGTCCCA ATTTACAATG TTGAACAATA TCTATCCAAG
TGTATAAATA GCATTGTAAA TCAGACCTAC AAACATATAG AGATTCTTCT GGTGAATGAC GGTAGTACGG
ATAATTCGGA AGAAATTTGT TTAGCATATG CGAAGAAAGA
TAGTCGCATT CGTTATTTTA AAAAAGAGAA CGGCGGGCTA TCAGATGCCC GTAATTATGG CATAAGTCGC
GCCAAGGGTG ACTACTTAGC TTTTATAGAC TCAGATGATT
TTATTCATTC GGAGTTCATC CAACGTTTAC ACGAAGCAAT TGAGAGAGAG AATGCCCTTG TGGCAGTTGC
TGGTTATGAT AGGGTAGATG CTTCGGGGCA TTTCTTAACA
GCAGAGCCGC TTCCTACAAA TCAGGCTGTT CTGAGCGGCA GGAATGTTTG TAAAAAGCTG CTAGAGGCGG
ATGGTCATCG CTTTGTGGTG GCCTGGAATA AACTCTATAA
AAAAGAACTA TTTGAAGATT TTCGATTTGA AAAGGGTAAG ATTCATGAAG ATGAATACTT CACTTATCGC
TTGCTCTATG AGTTAGAAAA AGTTGCAATA GTTAAGGAGT
GCTTGTACTA TTATGTTGAC CGAGAAAATA GTATCATAAC TTCTAGTATG ACTGACCATC GCTTCCATTG
CCTACTGGAA TTTCAAAATG AACGAATGGA CTTCTATGAA
AGTAGAGGAG ATAAAGAGCT CTTACTAGAG TGTTATCGTT CATTTTTAGC CTTTGCTGTT TTGTTTTTAG
GCAAATATAA TCATTGGTTG AGCAAACAGC AAAAGAAGCT
TCTCCAAACG CTATTTAGAA TTGTATATAA ACAATTGAAG CAAAATAAGC GACTTGCTTT ACTAATGAAT
GCTTATTATT TGGTAGGGTG TCTTCATCTT AATTTTAGTG
TCTTTCTGAA AACGGGGAAA GATAAAATTC AAGAAAGATT GAGAAGAAGT GAAAGTAGTA CTCGGTAAGA
ATGTTGTAAT AAATGGTTGA AAGAAAAGGG GATTAAAATG
AATCCAACAA ATAGTAGAAT AGCACTCTTT GATACGATTA AATGTATCAT GGTACTTTGT GTTATTTTTA
CACATCTGGA TTGGTCTGTT GAGCAGCGTC CATGGTTTAT
CTTTCCGTAT TTCGTTGACA TGGCTGTTCC AATTTTCNGT TGCTTCTGCC TATTTTCN

Fig. 3 cont.

ORF Z

SLDIDHMMEVMEASKSAAGSACPSPQAYQAAFEGAENIIVVTITGGLSGSFNAARVARDM
YIEEHPNVNIHLIDSLSASGEMDLLVHQINRLISAGLDFPQVVEAITHYREHSKLLFVLA
KVDNLVKNGRLSKLVGTVVGLLNIRMVGEASAEGKLELLQKARGHKKSVTAAFEEMKKAG
YDGGRIVMAHRNNAKFFQQFSELVKASFPTAVIDEVATSGLCSFYAEEGGLLMGYEVKA

Fig. 3 cont.

ORF Y

MKKYQVIIQDILTGIEEHRFKRGEKLPSIRQLREQYHCSKDTVQKAMLELKYQNKIYAVE
KSGYYILEDRDFQDHTCRAQSYRLSRITYEDFRICLKESLIGRENYLFNYYHQQEGLAEL
ISSVQSLLMDYHVYTKKDQLVITAGSQQALYILTQMETLAGKTEILIENPTYSRMIELIR
HQGIPYQTIERNLDGIDLEELESIFQTGKIKFFYTIPRLHNPLGSTYDIATKTAIVKLAK
QYDVYIIEDDYLADFDSSHSLPLHYLDTDNRVIYIKSFTPTLFPALRIGAISLPNQLRDI
FIKHKSLIDYDTNLIMQKALSLYIDNGMFARNTQHLHHIYHAQWNKIKDCLEKYALNIPY
RIPKGSVTFQLSKGILSPSIQHMFGKCYYFSGQKADFLQIFFEQDFADKLEQFVRYLNE

Fig. 3 cont.

ORF X

MKIIIPNAKEVNTNLENASFYLLSDRSKPVLDAISQFDVKKMAAFYKLNEAKAELEADRW
YRIRTGQAKTYPAWQLYDGLMYRYMDRRGIDSKEENYLRDHVRVATALYGLIHPFEFISP
HRLDFQGSLKIGNQSLKQYWRPYYDQEVGDDELILSLASSEFEQVFSPQIQKRLVKILFM
EEKAGQLKVHSTISKKGRGRLLSWLAKNNIQELSDIQDFKVDGFEYCTSESTANQLTFXR
SIKM

Fig. 3 cont.

CPS2A

MKKRSGRSKSSKFKLVNFALLGLYSITLCLFLVTMYRYNILDFRYLNYIVTLLLVGVAVL
AGLLMWRKKARIFTALLLVFSLVITSVGIYGMQEVVKFSTRLNSNSTFSEYEMSILVPAN
SDITDVRQLTSILAPAEYDQDNITALLDDISKMESTQLATSPGTSYLTAYQSMLNGESQA
MVFNGVFTNILENEDPGFSSKVKKIYSFKVTQTVETATKQVSGDSFNIYISGIDAYGPIS
TVSRSDVNIIMTVNRATHKILLTTTPRDSYVAFADGGQNQYDKLTHAGIYGVNASVHTLE
NFYGIDISNYVRLNFISFLQLIDLVGGIDVYNDQEFTSLHGNYHFPVGQVHLNSDQALGF
VRERYSLTGGDNDRGKNQEKVIAALIKKMSTPENLKNYQAILSGLEGSIQTDLSLETIMS
LVNTQLESGTQFTVESQALTGTGRSDLSSYAMPGSQLYMMEINQDSLEQSKAAIQSVLVE
K

Fig. 3 cont.

CPS2B

MNNQEVNAIEIDVLFLLKTIWRKKFLILLTAVLTAGLAFVYSSFLVTPQYDSTTRIYVVS
QNVEAGAGLTNQELQAGTYLAKDYREIILSQDVLTQVATELNLKESLKEKISVSIPVDTR
IVSISVRDADPNEAARIANSLRTFAVQKVVEVTKVSDVTTLEEAVPAEEPTTPNTKRNIL
LGLLAGGILATGLVLVMEVLDDRVKRPQDIEEVMGLTLLGIVPDSKKLK

Fig. 3 cont.

CPS2C

MAMLEIARTKREGVNKTEEYFNAIRTNIQLSGADIKVVGITSVKSNEGKSTTAASLAIAY
ARSGYKTVLVDADIRNSVMPGFFKPITKITGLTDYLAGTTDLSQGLCDTDIPNLTVIESG
KVSPNPTALLQSKNFENLLATLRRYYDYVIVDCPPLGLVIDAAIIAQKCDAMVAVVEAGN
VKCSSLKKVKEQLEQTGTPFLGVILNKYDIATEKYSEYGNYGKKA

Fig. 3 cont.

CPS2D

MIDIHSHIIFGVDDGPKTIEESLSLISEAYRQGVRYIVATSHRRKGMFETPEKIIMINFL
QLKEAVAEVYPEIRLCYGAELYYSKDILSKLEKKKVPTLNGSCYILLEFSTDTPWKEIQE
AVNEMTLLGLTPVLAHIERYDALAFQSERVEKLIDKGCYTQVNSNHVLKPALIGERAKEF
KKRTRYFLEQDLVHCVASDMHNLYSRPPFMREAYQLVKKEYGEDRAKALFKKNPLLILKN
QVQ

Fig. 3 cont.

CPS2E

MNIEIGYRQTKLALFDMIAVTISAILTSHIPNADLNRSGIFIIMMVHYFAFFISRMPVEF
EYRGNLIEFEKTFNYSIIFVIFLMAVSFMLENNFALSRRGAVYFTLINFVLVYLFNVIIK
QFKDSFLFSTTYQKKTILITTAELWENMQVLFESDILFQKNLVALVILGTEIDKINLPLP
LYYSVEEAIGFSTREVVDYVFINLPSEYFDLKQLVSDFELLGIDVGVDINSFGFTVLKNK
KIQMLGDHSIVTFSTNFYKPSHIWMKRLLDILGAVVGLIISGIVSILLIPIIRRDGGPAI
FAQKRVGQNGRIFTFYKFRSMFVDAEVRKKELMAQNQMQGGMFKMDNDPRITPIGHFIRK
TSLDELPQFYNVLIGDMSLVGTRPPTVDEFEKYTPSQKRRLSFKPGITGLWQVSGRSDIT
DFNEVVRLDLTYIDNWTIWSDIKILLKTVKVVLLREGGQ

Fig. 3 cont.

CPS2F

MRTVYIIGSKGIPAKYGGFETFVEKLTEYQKDKSINYFVACTRENSAKSDITGEVFEHNG
ATCFNIDVPNIGSAKAILYDIMALKKSIEIAKDRNDTSPIFYILACRIGPFIYLFKKQIE
SIGGQLFVNPDGHEWLREKWSYPVRQYWKFSESLMLKYADLLICDSKNIEKYIHEDYRKY
APETSYIAYGTDLDKSRLSPTDSVVREWYKEKEISENDYYLVVGRFVPENNYEVMIREFM
KSYSRKDFVLITNVEHNSFYEKLKKETGFDKDKRIKFVGTVYNQELLKYIRENAFAYFHG
HEVGGTNPSLLEALSSTKLNLLLDVGFNREVGEEGAKYWNKDNLHRVIDSCEQLSQEQIN
DMDSLSTKQVKERFSWDFIVDEYEKLFKG

Fig. 3 cont.

CPS2G

MKKILYLHAGAELYGADKVLLELIKGLDKNEFEAHVILPNDGVLVPALREVGAQVEVINY
PILRRKYFNPKGIFDYFISYHHYSKQIAQYAIENKVDIIHNNTTAVLEGIYLKRKLKLPL
LWHVHEIIVKPKFISDSINFLMGRFADKIVTVSQAVANHIKQSPHIKDDQISVIYNGVDN
KVFYQSDARSVRERFDIDEEALVIGMVGRVNAWKGQGDFLEAVAPILEQNPKAIAFIAGS
AFEGEEWRVVELEKKISQLKVSSQVXRMDYYANTTELYNMFDIFVLPSTNPDPLPTVVLK
AMACGKPVVGYRHGGVCEMVKEGVNGFLVTPNSPLNLSKVILQLSENINLRKKIGNNSIE
RQKEHFSLKSYVKNFSKVYTSLKVY

Fig. 3 cont.

CPS2H

MKIISFTMVNNESEIIESFIRYNYNFIDEMVIIDNGCTDNTMQIIFNLIKEGYKISVYDE
SLEAYNQYRLDNKYLTKIIAEKNPDLIIPLDADEFLTADSNPRKLLEQLDLEKIHYVNWQ
WFVMTKKDDINDSFIPRRMQYCFEKPVWHHSDGKPVTKCIISAKYYKKMNLKLSMGHHTV
FGNPNVRIEHHNDLKFAHYRAISQEQLIYKTICYTIRDIATMENNIETAQRTNQMALIES
GVDMWETAREASYSGYDCNVIHAPIDLSFCKENIVIKYNELSRETVAERVMKTGREMAVR
AYNVERKQKEKKFLKPIIFVLDGLKGDEYIHPNPSNHLTILTEMYNVRGLLTDNHQIKFL
KVNYRLIITPDFAKFLPHEFIVVPDTXDIEQVKSQYVGTGVDLSKIISLKEYRKEIGFIG
NLYALLGFVPNMLNRIYLYIQRNGIANTIIKIKSRL.

Fig. 3 cont.

CPS2I

MQADRRKTFGKMRIRINNLFFVAIAFMGIIISNSQVVLAIGKASVIQYLSYLVLILCIVN
DLLKNNKHIVVYKLGYLFLIIFLFTIGICQQILPITTKIYLSISMMIISVLATLPISLIK
DIDDFRRISNHLLFALFITSILGIKMGATMFTGAVEGIGFSQGFNGGLTHKNFFGITILM
GFVLTYLAYKYGSYKRTDRFILGLELFLILISNTRSVYLILLLFLFLVNLDKIKIEQRQW
STLKYISMLFCAIFLYYFFGFLITHSDSYAHRVNGLINFFEYYRNDWFHLMFGAADLAYG
DLTLDYAIRVRRVLGWNGTLEMPLLSIMLKNGFIGLVGYGIVLYKLYRNVRILKTDNIKT
IGKSVFIIVVLSATVENYIVNLSFVFMPICFCLLNSISTMESTINKQLQT

Fig. 3 cont.

CPS2J

MEKVSIIVPIFNTEKYLRECLDSIISQSYTNLEILLIDDGSSDSSTDICLEYAEQDGRIK
LFRLPNGGVSNARNYGIKNSTANYIMFVDSDDIVDGNIVESLYTCLKENDSDLSGGLLAT
FDGNYQESELQKCQIDLEEIKEVRDLGNENFPNHYMSGIFNSPCCKLYKNIYINQGFDTE
QWLGEDLLFNLNYLKNIKKVRYVNRNLYFARRSLQSTTNTFKYDVFIQLENLEEKTFDLF
VKIFGGQYEFSVFKETLQWHIIYYSLLMFKNGDESLPKKLHIFKYLYNRHSLDTLSIKRT
SSVFKRICKLIVANNLFKIFLNTLIREEKNND

Fig. 3 cont.

CPS2K

MINISIIVPIYNVEQYLSKCINSIVNQTYKHIEILLVNDGSTDNSEEICLAYAKKDSRIR
YFKKENGGLSDARNYGISRAKGDYLAFIDSDDFIHSEFIQRLHEAIERENALVAVAGYDR
VDASGHFLTAEPLPTNQAVLSGRNVCKKLLEADGHRFVVAWNKLYKKELFEDFRFEKGKI
HEDEYFTYRLLYELEKVAIVKECLYYYVDRENSIITSSMTDHRFHCLLEFQNERMDFYES
RGDKELLLECYRSFLAFAVLFLGKYNHWLSKQQKKLLQTLFRIVYKQLKQNKRLALLMNA
YYLVGCLHLNFSVFLKTGKDKIQERLRRSESSTR.

Fig. 3 cont.

ATCGCCAAAC GAAATTGGCA TTATTTGATA TGATAGCAGT TGCAATTTCT GCAATCTTAA CAAGTCATAT
ACCAAATGCT GATTTAAATC GTTCTGGAAT TTTTATCATA
ATGATGGTTC ATTATTTTGC ATTTTTTATA TCTCGTATGC CAGTTGAATT TGAGTATAGA GGTAATCTGA
TAGAGTTTGA AAAAACATTT AACTATAGTA TAATATTTGC
AATTTTTCTT ACGGCAGTAT CATTTTTGTT GGAGAATAAT TTCGCACTTT CAAGACGTGG TGCCGTGTAT
TTCACATTAA TAAACTTCGT TTTGGTATAC CTATTTAACG
TAATTATTAA GCAGTTTAAG GATAGCTTTC TATTTTCGAC AATCTATCAA AAAAAGACGA TTCTAATTAC
AACGGCTGAA CGATGGGAAA ATATGCAAGT TTTATTTGAA
TCACATAAAC AAATTCAAAA AAATCTTGTT GCATTGGTAG TTTTAGGTAC AGAAATAGAT AAAATTAATT
TATCATTACC GCTCTATTAT TCTGTGGAAG AAGCTATAGA
GTTTTCAACA AGGGAAGTGG TCGACCACGT CTTTATAAAT CTACCAAGTG AGTTTTTAGA CGTAAAGCAA
TTCGTTTCAG ATTTTGAGTT GTTAGGTATT GATGTAAGCG
TTGATATTAA TTCATTCGGT TTTACTGCGT TGAAAAACAA AAAAATCCAA CTGCTAGGTG ACCATAGCAT
TGTAACTTTT TCCACAAATT TTTATAAGCC TAGTCATATC
ATGATGAAAC GACTTTTGGA TATACTCGGA GCGGTAGTCG GGTTAATTAT TTGTGGTATA GTTTCTATTT
TGTTAGTTCC AATTATTCGT AGAGATGGTG GACCGGCTAT
TTTTGCTCAG AAACGAGTTG GACAGAATGG ACGCATATTT ACATTCTACA AGTTTCGATC GATGTATGTT
GATGCTGAGG AGCGCAAAAA AGACTTGCTC AGCCAAAACC
AGATGCAAGG GTGGGTATGT TTTAAAATGG GAAAAACGAT CCTAGAATTA CTCCAATTGG ACATTTCATA
CGCAAAAACA AGTTTAGACG AGTTACCACA GTTTTATAAT
GTTTTAATTG GCGATATGAG TCTAGTTGGT ACACGTCCAC CTACAGTTGA TGAATTTGAA AAATATACTC
CTGGTCAAAA GAGACGATTG AGTTTTAAAC CAGGGATTAC
AGGTCTCTGG CAGGTTAGTG GTCGTAGTAA TATCACAGAC TTCGACGACG TAGTTCGGTT GGACTTAGCA
TACATTGATA ATTGGACTAT CTGGTCAGAT ATTAAAATTT
TATTAAAGAC AGTGAAAGTT GTATTGTTGA GAGAGGGAAG TAAGTAAAAG TATATGAAAG TTTGTTTGGT
CGGTTCTTCA GGGGGACATT TGACTCACTT GTATTTGTTA
AAACCGTTTT GGAAGGAAGA AGAACGTTTT TGGGTAACAT TTGATAAAGA GGATGCAAGA AGTCTTTTGA
AGAATGAAAA AATGTATCCA TGTTACTTTC CAACAAATCG
CAATCTCATT AATTTAGTGA AAAATACTTT CTTAGCTTTC AAAATTTTAC GTGATGAGAA ACCAGATGTT
ATTATTTCAT CTGGTGCGGC CGTTGCTGTC CCCTTCTTTT
ACATCGGAAA ACTATTTGGA GCAAAGACGA TTTATATTGA AGTATTTGAT CGAGTTAATA AATCTACATT
AACTGGAAAA CTAGTTTATC CCGTAACAGA TATTTTTATT
GTTCAGTGGG AAGAAATGAA GAAGGTATAT CCTAAATCTA TTAACTTGGG GAGTATTTTT TAATGATTTT
TGTAACAGTA GGAACTCATG AACAACAGTT TAATCGATTG
ATAAAAGAGA TTGATTTATT GAAAAAAAAT GGAAGTATAA CCGACGAAAT ATTTATTCAA ACAGGATATT
CTGACTATAT TCCAGAATAT TGCAAGTATA AAAAATTTCT
CAGTTACAAA GAAATGGAAC AATATATTAA CAAATCAGAA GTAGTTATTT GCCACGGAGG CCCCGCTACT
TTTATGAATT CATTATCCAA AGGAAAAAAA CAATTATTGT
TTCCTAGACA AAAAAAGTAT GGTGAACATG TAAATGATCA TCAAGTAGAG TTTGTAAGAA GAATTTTACA
AGATAATAAT ATTTTATTTA TAGAAAATAT AGATGATTTG
TTTGAAAAAA TTATTGAAGT TTCTAAGCAA ACTAACTTTA CATCAAATAA TAATTTTTTT TGTGAAAGAT
TAAAACAAAT AGTTGAAAAA TTTAATGAGG ATCAAGAAAA
TGAATAATAA AAAAGATGCA TATTTGATAA TGGCTTATCA TAATTTTTCT CAGATTTTAC TGGAGAGGGA
TACAGATATT ATCATCTTCT CTCAGGAGAA TGCACACCAT
TAGTTCCTTC AGAATACCTG TATAATTATT TTAAATATTC TCAGGATTTA TATGTTGAAT TTACAAAAGA
TGAGCAAAAA TATAAAGAAA ATAGGATATA TGAACGAGTT
AAATGTTACA GATTATTTCC TAATATATCA GAAAAAACTA TTGATAATGT ACTGTTTAGA ATTTTATTAA
GAATGTATCG AGCTTTTGAA TACTATTTAC AAAGATTGTT
GTTTATTGAT AGAATAAAAA ACATGGTCTA AGAATAAGAT TTGGTTCTAA TTGGGTTTCG CTTCCACATG
ATTTTGTGGC AATTCTTTTA TCAAATGAAA ACGAAACAGC
TTATTTATTT AAGTAATCTA AATGTCCAGA TGAACTATTT ATACAGACAA TTATAGAAAA ATATGAATTT
TCAAATAGAT TATCTAAATA TGGAAATTTA AGATATATAA
AGTGGAAAAA ATCAACATCT TCTCCTATTG TCTTTACAGA TGATTCTATT GATGAATTGC TAAATGCAAG
AAATTTAGGT TTTTTATTTG CTAGAAAGTT AAAAATAGAA
AATAAATCTA AATTTAAAGA AATTATTACT AAAAAATAAA ATAGTTGATT TTGTGAGAGT AATGTATGTT
TAAATTATTT AAATATGACC CGGAATATTT TATTTTTAAG
TACTTCTGGT TGATTATTTT TATTCCAGAG CAAAAGTATG TATTTTTATT AATTTTTATG AATTTAATTT
TATTTCATAT AAAATTTTTG AAAACTAAGC TAATATTAAA
AAATGAAATT TTATTGTTTT TATTATGGTC TATATTATGT TTTGTTTCAG TAGTCACAAG TATGTTTGTT
GAAATAAATT TTGAAAGATT ATTTGCAGAT TTTACTGCTC
CCATAATTTG GATTATTGCA ATAATGTATT ATAATTTGTA TTCATTTATA AATATTGATT ATAAAAAATT

Fig. 4

AAAAAATAGT ATCTTTTTTA GTTTTTTAGT TTTATTAGGT
ATATCTGCAT TGTATATTAT TCAAAATGGG AAAGATATTG TATTTTTAGA CAGACACCTT ATAGGACTAG
ACTATCTTAT AACAGGCGTC AAAACAAGGT TGGTTGGCTT
TATGAACTAT CCTACGTTAA ATACCACTAC AATTATAGTT TCAATTCCGT TAATCTTTGC ACTTATAAAA
AATAAAATGC AACAATTTTT TTTCTTGTGT CTTGCTTTTA
TACCGATCTA TTTAAGTGGA TCGAGAATTG GTAGTTTATC GCTAGCAATA TTAATTATAT GCTTGTTATG
GAGATATATA GGTGGAAAAT TTGCTTGGAT AAAAAAGCTA
ATAGTAATAT TTGTAATACT ACTTATTATT TTAAATACTG AATTGCTTTA CCATGAAATT TTGGCTGTTT
ATAATTCTAG AGAATCAAGT AACGAAGCTA GATTTATTAT
TTATCAAGGA AGTATTGATA AAGTATTAGA AAACAATATT TTATTTGGAT ATGGAATATC CGAATATTCA
GTTACGGGAA CTTGGCTCGG AAGTCATTCA GGCTATATAT
CATTTTTTTA TAAATCAGGA ATAGTTGGGT TGATTTTACT GATGTTTTCT TTTTTTTATG TTATAAAAAA
AAGTTATGGA GTTAATGGGG AAACAGCACT ATTTTATTTT
ACATCATTAG CCATATTTTT CATATATGAA ACAATAGATC CGATTATTAT TATATTAGTA CTATTCTTTT
CTTCAATAGG TATTTGGAAT AATATAAATT TTAAAAAGGA
TATGGAGACA AAAAATGAAT GATTTAATTT CAGTTATTGT ACCAATTTAT AATGTCCAAG ATTATCTTGA
TAAATGTATT AACAGTATTA TTAACCAAAC ATATACTAAT
TTAGAGGTTA TTCTCGTAAA TGATGGAAGT ACTGATGATT CTGAGAAAAT TTGCTTAAAC TATATGAAGA
ACGATGGAAG AATTAAATAT TACAAGAAAA TTAATGGCGG
TCTAGCAGAT GCTCGAAATT TCGGACTAGA ACATGCAACA GGTAAATATA TTGCTTTTGT CGATTCTGAT
GACTATATAG AAGTTGCAAT GTTCGAGAGA ATGCATGATA
ATATAACTGA GTATAATGCC GATATAGCAG AGATAGATTT TTGTTTAGTA GACGAAAACG GGTATACAAA
GAAAAAAAGA AATAGTAATT TTCATGTCTT AACGAGAGAA
GAGACTGTAA AAGAATTTTT GTCAGGATCT AATATAGAAA ATAATGTTTG GTGCAAGCTT TATTCACGAG
ATATTATAAA AGATATAAAA TTCCAAATTA ATAATAGAAG
TATTGGTGAG GATTTGCTTT TTAATTTGGA GGTCTTGAAC AATGTAACAC GTGTAGTAGT TGATACTAGA
GAATATTATT ATAATTATGT CATTCGTAAC AGTTCGCTTA
TTAATCAGAA ATTCTCTATA AATAATATTG ATTTAGTCAC AAGATTGGAG AATTACCCCT TTAAGTTAAA
AAGAGAGTTT AGTCATTATT TTGATGCAAA AGTTATTAAA
GAGAAGGTTA AATGTTTAAA CAAAATGTAT TCAACAGATT GTTTGGATAA TGAGTTCTTG CCAATATTAG
AGTCTTATCG AAAAGAAATA CGTAGATATC CATTTATTAA
AGCGAAAAGA TATTTATCAA GAAAGCATTT AGTTACGTTG TATTTGATGA AATTTTCGCC TAAACTATAT
GTAATGTTAT ATAAGAAATT TCAAAAGCAG TAGAGGTAAA
AATGGATAAA ATTAGTGTTA TTGTTCCAGT TTATAATGTA GATAAATATT TAAGTAGTTG TATAGAAAGC
ATTATTAATC AAAATTATAA AAATATAGAA ATATTATTGA
TAGATGATGG CTCTGTAGAT GATTCTGCTA AAATATGCAA GGAATATGCA GAAAAAGATA AAAGAGTAAA
AATTTTTTTC ACTAATCATA GTGGAGTATC AAATGCTAGA
AATCATGGAA TAAAGCGGAG TACAGCTGAA TATATTATGT TTGTTGACTC TGATGATGTT GTTGATAGTA
GATTAGTAGA AAAATTATAT TTTAATATTA TAAAAAGTAG
AAGTGATTTA TCTGGTTGTT TGTACGCTAC TTTTTCAGAA AATATAAATA ATTTTGAAGT GAATAATCCA
AATATTGATT TTGAAGCAAT TAATACCGTG CAGGACATGG
GAGAAAAAAA TTTTATGAAT TTGTATATAA ATAATATTTT TTCTACTCCT GTTTGTAAAC TATATAAGAA
AAGATACATA ACAGATCTTT TTCAAGAGAA TCAATGGTTA
GGAGAAGATT TACTTTTTAA TCTGCATTAT TTAAAGAATA TAGATAGAGT TAGTTATTTG ACTGAACATC
TTTATTTTTA TAGGAGAGGT ATACTAAGTA CAGTAAATTC
TTTTAAAGAA GGTGTGTTTT TGCAATTGGA AAATTTGCAA AAACAAGTGA TAGTATTGTT TAAGCAAATA
TATGGTGAGG ATTTTGACGT ATCAATTGTT AAAGATACTA
TACGTTGGCA AGTATTTTAT TATAGCTTAC TAATGTTTAA ATACGGAAAA CAGTCTATTT TTGACAAATT
TTTAATTTTT AGAAATCTTT ATAAAAAATA TTATTTTAAC
TTGTTAAAAG TATCTAACAA AAATTCTTTG TCTAAAAATT TTTGTATAAG AATTGTTTCG AACAAAGTTT
TTAAAAAAAT ATTATGGTTA TAATAGGAAG ATATCATGGA
TACTATTAGT AAAATTTCTA TAATTGTACC TATATATAAT GTAGAAAAAT ATTTATCTAA ATGTATAGAT
AGCATTGTAA ATCAGACCTA CAAACATATA GAGATTCTTC
TGGTGAATGA CGGTAGTACG GATAATTCGG AAGAAATTTG TTTAGCATAT GCGAAGAAAG ATAGTCGCAT
TCGTTATTTT AAAAAAGAGA ACGGCGGGCT ATCAGATGCC
CGTAATTATG GCATAAGTCG CGCCAAGGGT GACTACTTAG CTTTTATAGA CTCAGATGAT TTTATTCATT
CGGAGTTCAT CCAACGTTTA CACGAAGCAA TTGAGAGAGA
GAATGCCCTT GTGGCAGTTG CTGGTTATGA TAGGGTAGAT GCTTCGGGGC ATTTCTTAAC AGCAGAGCCG
CTTCCTACAA ATCAGGCTGT TCTGAGCGGC AGGAATGTTT
GTAAAAAGCT GCTAGAGGCG GATGGTCATC GCTTTGTGGT GGCCTGTAAT AAACTCTATA AAAAAGAACT
ATTTGAAGAT TTTCGATTTG AAAAGGGTAA GATTCATGAA

Fig. 4 cont.

GATGAATACT TCACTTATCG CTTGCTCTAT GAGTTAGAAA AAGTTGCAAT AGTTAAGGAG TGCTTGTACT
ATTATGTTGA CCGAGAAAAT AGTATCACAA CTTCTAGCAT
GACTGACCAT CGCTTCCATT GCCTACTGGA ATTTCAAAAT GAACGAATGG ACTTCTATGA AAGTAGAGGA
GATAAAGAGC TCTTACTAGA GTGTTATCGT TCATTTTTAG
CCTTTGCTGT TTTGTTTTTA GGCAAATATA ATCATTGGTT GAGCAAACAG CAAAAGAAGC TT

Fig. 4 cont.

CPS1E

RQTKLALFDMIAVAISAILTSHIPNADLNRSGIFIIMMVHYFAFFISRMPVEFEYRGNLI
EFEKTFNYSIIFAIFLTAVSFLLENNFALSRRGAVYFTLINFVLVYLFNVIIKQFKDSFL
FSTIYQKKTILITTAERWENMQVLFESHKQIQKNLVALVVLGTEIDKINLSLPLYYSVEE
AIEFSTREVVDHVFINLPSEFLDVKQFVSDFELLGIDVSVDINSFGFTALKNKKIQLLGD
HSIVTFSTNFYKPSHIMMKRLLDILGAVVGLIICGIVSILLVPIIRRDGGPAIFAQKRVG
QNGRIFTFYKFRSMYVDAEERKKDLLSQNQMQGWVCFKMGKTILELLQLDISYAKTSLDE
LPQFYNVLIGDMSLVGTRPPTVDEFEKYTPGQKRRLSFKPGITGLWQVSGRSNITDFDDV
VRLDLAYIDNWTIWSDIKILLKTVKVVLLREGSK

Fig. 4 cont.

CPS1F

MKVCLVGSSGGHLTHLYLLKPFWKEEERFWVTFDKEDARSLLKNEKMYPCYFPTNRNLIN
LVKNTFLAFKILRDEKPDVIISSGAAVAVPFFYIGKLFGAKTIYIEVFDRVNKSTLTGKL
VYPVTDIFIVQWEEMKKVYPKSINLGSIF

Fig. 4 cont.

CPS1G

MIFVTVGTHEQQFNRLIKEIDLLKKNGSITDEIFIQTGYSDYIPEYCKYKKFLSYKEMEQ
YINKSEVVICHGGPATFMNSLSKGKKQLLFPRQKKYGEHVNDHQVEFVRRILQDNNILFI
ENIDDLFEKIIEVSKQTNFTSNNNFFCERLKQIVEKFNEDQENE

Fig. 4 cont.

CPS1H

MFKLFKYDPEYFIFKYFWLIIFIPEQKYVFLLIFMNLILFHIKFLKTKLILKNEILLFLL
WSILCFVSVVTSMFVEINFERLFADFTAPIIWIIAIMYYNLYSFINIDYKKLKNSIFFSF
LVLLGISALYIIQNGKDIVFLDRHLIGLDYLITGVKTRLVGFMNYPTLNTTTIIVSIPLI
FALIKNKMQQFFFLCLAFIPIYLSGSRIGSLSPLAILIICLLWRYIGGKFAWIKKLIVIF
VILLIIILNTELLYHEILAVYNSRESSNEARFIIYQGSIDKVLENNILFGYGISEYSVTGT
WLGSHSGYISFFYKSGIVGLILLMFSFFYVIKKSYGVNGETALFYFTSLAIFFIYETIDP
IIIILVLFFSSIGIWNNINFKKDMETKNE

Fig. 4 cont.

CPS1I

MNDLISVIVPIYNVQDYLDKCINSIINQTYTNLEVILVNDGSTDDSEKICLNYMKNDGRI
KYYKKINGGLADARNFGLEHATGKYIAFVDSDDYIEVAMFERMHDNITEYNADIAEIDFC
LVDENGYTKKKRNSNFHVLTREETVKEFLSGSNIENNVWCKLYSRDIIKDIKFQINNRSI
GEDLLFNLEVLNNVTRVVVDTREYYYNYVIRNSSLINQKFSINNIDLVTRLENYPFKLKR
EFSHYFDAKVIKEKVKCLNKMYSTDCLDNEFLPILESYRKEIRRYPFIKAKRYLSRKHLV
TLYLMKFSPKLYVMLYKKFQKQ

Fig. 4 cont.

CPS1F

```
MDKISVIVPVYNVDKYLSSCIESIINQNYKNIEILLIDDGSVDDSAKICKEYEKDKRVKI
FFTNHSGVSNARNHGIKRSTAEYIMFVDSDDVVDSRLVEKLYFNIIKSRSDLSGCLYATF
SENINNFEVNNPNIDFEAINTVQDMGEKNFMNLXXNNIFSTPVCXLYQKRYITDLFQENQ
WLGEDLLFNLHYLKNIDRVSYLTEHLYFYRRGILSTVNSFKEGVFLQLENLQKQVIVLFK
QIYGEDFDVSIVKDTIRWQVFYYSLLMFKYGKQSIFDKFLIFRNLYKKYYFNLLKVSNKN
SLSKNFCIRIVSNKVFKKILWL
```

Fig. 4 cont.

CPS1K

MDTISKISIIVPIYNVEKYLSKCIDSIVNQTYKHIEILLVNDGSTDNSEEICLAYAKKDS
RIRYFKKENGGLSDARNYGISRAKGDYLAFIDSDDFIHSEFIQRLHEAIERENALVAVAG
YDRVDASGHFLTAEPLPTNQAVLSGRNVCKKLLEADGHRFVVACNKLYKKELFEDFRFEK
GKIHEDEYFTYRLLYELEKVAIVKECLYYYVDRENSITTSSMTDHRFHCLLEFQNERMDF
YESRGDKELLLECYRSFLAFAVLFLGKYNHWLSKQQKK

Fig. 4 cont.

CPS9

AAGCTTATCG TCAAGGTGTT CGCTATATCG TGGCGACATC TCATAGACGA AAAGGGATGT TTGAAACACC
AGAAAAAGTT ATCATGACTA ACTTTCTTCA ATTTAAAGAC
GCAGTAGCAG AAGTTTATCC TGAAATACGA TTGTGCTATG GTGCTGAATT GTATTATAGT AAAGATATAT
TAAGCAAACT TGAAAAAAAG AAAGTACCCA CACTTAATGG
CTCGCGCTAT ATTCTTTTGG AGTTCAGTAG TGATACTCCT TGGAAAGAGA TTCAAGAAGC AGTGAACGAA
GTGACGCTAC TTGGGCTAAC TCCCGTACTT GCCCATATAG
AACGATATGA CGCCCTAGCG TTTCATGCAG AGAGAGTAGA AGAGTTAATT GACAAGGGAT GCTATACTCA
GGTAAATAGT AATCATGTGC TGAAGCCCAC TTTAATTGGT
GATCGAGCAA AAGAATTTAA AAAACGTACT CGGTATTTTT TAGAGCAGGA TTTAGTACAT TGTGTTGCTA
GCGATATGCA TAATTTATCT AGTAGACCTC CGTTTATGAG
GGAGGCTTAT AAGTTGCTAA CAGAGGAATT TGGCAAAGAT AAAGCGAAAG CGTTGCTAAA AAAGAATCCT
CTTATGCTAT TAAAAAACCA GGCGATTTAA ACTGGTTACT
CTAGATTGTG GAGAGAAAAA TGGATTTAGG AACTGTTACT GATAAACTGT TAGAACGCAA CAGTAAACGA
TTGATACTCG TGTGCATGGA TACGTGTCTT CTTATAGTTT
CCATGATTTT GAGCAGACTG TTTTTGGATG TTATTATTGA CATACCAGAT GAACGCTTCA TTCTTGCAGT
TTTATTCGTA TCAATTTTAT ATTTGATTCT ATCGTTTAGA
TTAAAAGTCT TTTCATTAAT TACGCGTTAC ACAGGGTATC AGAGTTATGT AAAAATAGGA CTTAGTTTAA
TATCTGCGCA TTCATTGTTT TTAATTATCT CAATGGTGTT
GTGGCAGGCT TTTAGTTATC GTTTCATCTT AGTATCCTTA TTTTTGTCGT ATGTAATGCT CATTACTCCG
AGGATTGTTT GGAAAGTCTT ACATGAGACG AGAAAAAATG
CTATCCGTAA GAAGGATAGC CCACTAAGAA TCTTAGTAGT AGGTGCTGGA GATGGTGGTA ATATTTTTAT
CAATACTGTC .AAAGATCGAA AATTGAATTT TGAAATTGTC
GGTATCGTTG ATCGTGATCC AAATAAACTT GGAACATTTA TCCGTACGGC TAAAGTTTTA GGAAACCGTA
ATGATATTCC ACGACTGGTA GAGGAATTAG CTGTTGACCA
AGTGACGATT GCCATCCCTT CTTTAAATGG TAAGGAGCGA GAGAAGATTG TTGAAATCTG TAACACTACA
GGAGTGACCG TCAATAATAT GCCGAGTATT GAAGACATTA
TGGCGGGGAA CATGTCTGTC AGTGCCTTTC AGGAAATTGA CGTAGCAGAC CTTCTTGGTC GACCAGAGGT
TGTTTTGGAT CAGGATGAAT TGAATCAGTT TTTCCAAGGG
AAAACAATCC ·TTGTCACAGG AGCAGGTGGC TCTATCGGTT CAGAGCTATG TCGTCAAATT GCTAAGTTTA
CGCCTAAACG CTTGTTGTTG CTTGGACATG GAGAAAATTC
AATCTATCTC ATTCATCGAG AGTTACTGGA AAAGTACCAA GGTAAGATTG AGTTGGTCCC TCTCATTGCA
GATATTCAAG ATAGAGAATT GATTTTTAGC ATAATGGCTG
AATATCAACC CGATGTTGTT TATCATGCTG CAGCACATAA GCATGTTCCT TTGATGGAAT ATAATCCACA
TGAAGCAGTG AAGAATAATA TTTTTGGAAC GAAGAATGTG
GCTGAGGCGG CTAAAACTGC AAAGGTTGCC AAATTTGTTA TGGTTTCAAC AGATAAAGCT GTTAATCCAC
CAAATGTCAT GGGAGCGACT AAACGTGTTG CAGAAATGAT
TGTTACAGGT TTAAACGAGC CAGGTCAGAC TCAATTTGCG GCAGTCCGGT TTGGGAATGT TCTAGGTAGT
CGTGGAAGTG TTGTTCCGCT ATTCAAAGAG CAAATTAGAA
AAGGTGGACC TGTTACGGTT ACCGACTTTA GGATGACTCG TTATTTCATG ACGATTCCTG AGGCAAGTCG
TTTGGTTATC CAAGCTGGAC ATTTGGCAAA AGGTGGAGAA
ATATTTGTCT TGGATATGGG CGAGCCAGTA CAAATCCTGG AATTGGCAAG AAAAGTTATC TTGTTAAGTG
GACACACAGA GGAAGAAATC GGGATTGTAG AATCTGGAAT
CAGACCAGGC GAGAAACTCT ACGAGGAATT ATTATCAACA GAAGAACGTG TCAGCGAACA GATTCATGAA
AAAATATTTG TGGGTCGCGT TACAAATAAG CAGTCGGACA
TTGTCAATTC ATTTATCAAT GGATTACTCC AAAAAGATAG AAATGAATTA AAAAATATGT TGATTGAATT
TGCAAAACAA GAATAAGAAA GTAAAAAATA TTTTTACTTT
CCTAGAGTTT AAACGATGTT TAAGTTCTAG GAAGGTTAGA ATACCTAATT AACAACAATA TTACTATTTA
TTAAGAGTCA GATAATAGCA ACTAAGTGCT ACAAACTATC
TTTATAATAA GTATATTTGG TCAAAAGGGA GATGTGAAAT GTATCCAATT TGTAAACGTA TTTTAGCAAT
TATTATCTCA GGGATTGCTA TTGTTGTTCT GAGTCCAATT
TTATTATTGA TTGCATTGGC AATTAAATTA GATTCTAAAG GTCCGGTATT ATTTAAACAA AAGCGGGTTG
GTAAAAACAA GTCATACTTT ATGATTTATA AATTCCGTTC
TATGTACGTT GACGCACCAA GTGATATGCC GACTCATCTA TTAAAGGATC CTAAGGCGAT GATTACCAAG
GTGGGCGCGT TTCTCAGAAA AACAAGTTTA GATGAACTGC
CACAGCTTTT TAATATTTTT AAAGGTGAAA TGGCGATTGT TGGTCCACGC CCAGCCTTAT GGAATCAATA
TGACTTAATT GAAGAGCGAG ATAAATATGG TGCAAATGAT
ATTCGTCCTG GACTAACCGG TTGGGCTCAA ATTAATGGTC GTGATGAATT GGAAATTGAT GAAAAGTCAA
AATTAGATGG ATATTATGTT CAAAATATGA GTCTAGGTTT
GGATATTAAA TGTTTCTTAG GTACATTCCT CAGTGTAGCC AGAAGCGAAG GTGTTGTTGA AGGTGGAACA
GGGCAGAAAG GAAAAGGATG AAATTTTCAG TATTAATGTC
GGTCTATGAG AAAGAAAAAC CAGAGTTTCT TAGGGAATCT TTGGAAAGCA TCCTTGTCAA TCAAACAATG

Fig. 5

ATTCCAACGG AGGTTGTCTT GGTAGAGGAT GGGCCACTCA
ATCAGAGCTT ATATAGTATT TTAGAAGAAT TTAAAAGTCG ATTTTCATTT TTTAAAACGA TAGCCTTGGA
AAAGAATTCG GGTTTAGGAA TTGCACTGAA TGAAGGTTTG
AAACATTGTA ATTATGAGTG GGTTTGCACG AAATGGATTC TGATGATGTT GCATATACAT ACACGTTTTG
AAAAGCAAGT TAACTTTATA AAACAAAACC CGACTATAGA
TATTGAGATA GATGAGTTCT TAAATTCTAC TAGTGAAATA GTTTCTCATA AAAATGTTCC AACCCAGCAC
GATGAAATAT TAAAGATGGC AAGGCGGGAG AAATCCATGT
GCCACATGAC TGTAATGTTT AAAAAGAAAA GTGTCGAGAG AGCAGGGGGG TATCAAACAC TTCCGTACGT
AGAAGATTAT TTCCTTTGGG TGCGCATGAT TGCTTCAGGA
TCGAAATTTG CAAACATTGA TGAAACACTA GTTCTTGCAC GTGTTGGAAA TGGGATGTTC AATAGGAGGG
GGAACAGAGA ACAAATTAAC AGTTGGACAT TACTAATTGA
ATTTATGTTA GCTCAAGGAA TTGTTACACC ACTAGATGTA TTTATTAATC AAATTTACAT TAGGGTCTTT
GTTTATATGC CAACTTGGAT AAAGAAACTC ATTTATGGAA
AAATCTTAAG GAAATAGTAT GATTACAGTA TTGATGGCTA CATATAATGG AAGCCCATTT ATAATAAAAC
AGTTAGATTC AATTCGAAAT CAAAGTGTAT CAGCAGACAA
AGTTATTATT TGGGATGATT GCTCGACAGA TGATACAATA AAAATAATAA AAGATTATAT AAAAAAATAT
TCTTTGGATT CATGGGTTGT CTCTCAAAAT AAATCTAATC
AGGGGCATTA TCAAACATTT ATAAATTTGA CAAAGTTAGT TCAGGAAGGA ATAGTCTTTT TTTCAGATCA
AGATGATATT TGGGACTGTC ATAAAATTGA GACAATGCTT
CCAATCTTTG ACAGAGAAAA TGTATCAATG GTGTTTTGCA AATCCAGATT GATTGATGAA AACGGAAATA
TTATCAGTAG CCCAGATACT TCGGATAGAA TCAATACGTA
CTCTCTAGA

Fig 5 cont.

CPS9D

AYRQGVRYIVATSHRRKGMFETPEKVIMTNFLQFKDAVAEVYPEIRLCYGAELYYSKDIL
SKLEKKKVPTLNGSRYILLEFSSDTPWKEIQEAVNEVTLLGLTPVLAHIERYDALAFHAE
RVEELIDKGCYTQVNSNHVLKPTLIGDRAKEFKKRTRYFLEQDLVHCVASDMHNLSSRPP
FMREAYKLLTEEFGKDKAKALLKKNPLMLLKNQAI.

Fig. 5 cont.

CPS9E

MDLGTVTDKLLERNSKRLILVCMDTCLLIVSMILSRLFLDVIIDIPDERFILAVLFVSIL
YLILSFRLKVFSLITRYTGYQSYVKIGLSLISAHSLFLIISMVLWQAFSYRFILVSLFLS
YVMLITPRIVWKVLHETRKNAIRKKDSPLRILVVGAGDGGNIFINTVKDRKLNFEIVGIV
DRDPNKLGTFIRTAKVLGNRNDIPRLVEELAVDQVTIAIPSLNGKEREKIVEICNTTGVT
VNNMPSIEDIMAGNMSVSAFQEIDVADLLGRPEVVLDQDELNQFFQGKTILVTGAGGSIG
SELCRQIAKFTPKRLLLLGHGENSIYLIHRELLEKYQGKIELVPLIADIQDRELIFSIMA
EYQPDVVYHAAAHKHVPLMEYNPHEAVKNNIFGTKNVAEAAKTAKVAKFVMVSTDKAVNP
PNVMGATKRVAEMIVTGLNEPGQTQFAAVRFGNVLGSRGSVVPLFKEQIRKGGPVTVTDF
RMTRYFMTIPEASRLVIQAGHLAKGGEIFVLDMGEPVQILELARKVILLSGHTEEEIGIV
ESGIRPGEKLYEELLSTEERVSEQIHEKIFVGRVTNKQSDIVNSFINGLLQKDRNELKNM
LIEFAKQE

Fig. 5 cont.

CPS9F

MYPICKRILAIIISGIAIVVLSPILLLIALAIKLDSKGPVLFKQKRVGKNKSYFMIYKFR
SMYVDAPSDMPTHLLKDPKAMITKVGAFLRKTSLDELPQLFNIFKGEMAIVGPRPALWNQ
YDLIEERDKYGANDIRPGLTGWAQINGRDELEIDEKSKLDGYYVQNMSLGLDIKCFLGTF
LSVARSEGVVEGGTGQKGKG

Fig. 5 cont.

CPS9G

MKFSVLMSVYEKEKPEFLRESLESILVNQTMIPTEVVLVEDGPLNQSLYSILEEFKSRFS
FFKTIALEKNSGLGIALNEGLKHCNYEWVCTKWILMMLHIHTRFEKQVNFIKQNPTIDIE
IDEFLNSTSEIVSHKNVPTQHDEILKMARREKSMCHMTVMFKKKSVERAGGYQTLPYVED
YFLWVRMIASGSKFANIDETLVLARVGNGMFNRRGNREQINSWTLLIEFMLAQGIVTPLD
VFINQIYIRVFVYMPTWIKKLIYGKILRK

Fig. 5 cont.

CPS9H

MITVLMATYNGSPFIIKQLDSIRNQSVSADKVIIWDDCSTDDTIKIIKDYIKKYSLDSWV
VSQNKSNQGHYQTFINLTKLVQEGIVFFSDQDDIWDCHKIETMLPIFDRENVSMVFCKSR
LIDENGNIISSPDTSDRINTYSL

Fig. 5 cont.